



Business Intelligence em Dádivas de Sangue

FILOMENA FILIPA ARRÁTEL MATEUS

Outubro de 2019

Business Intelligence em Dádivas de Sangue

Filomena Filipa Arrátel Mateus

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas de Informação e Conhecimento**

Orientador: Paulo Oliveira

Júri:

Presidente:

Vogais:

Porto, outubro 2019

Dedicatória

Dedico esta dissertação aos meus pais que têm acompanhado todas as minhas decisões. Sem eles nada seria.

Resumo

A organização que forneceu os dados para se elaborar esta dissertação é responsável pela gestão e monitorização de Dádivas de Sangue a nível nacional. Esta armazena dados referentes aos dadores, às colheitas que os dadores realizam e ao resultado das análises de sangue.

O conjunto destas informações torna-se útil para os profissionais de saúde, já que estes são responsáveis pelas melhorias relacionadas com a prestação de cuidados de saúde. Assim, é de extrema importância a forma como a informação é disponibilizada, pois para estes interessa obter os dados o mais rápido possível para atuarem em conformidade.

Desenvolveu-se um conjunto de indicadores acerca dos dadores, colheitas e resultado de análises de sangue, de forma a retirar conclusões acerca do aumento/diminuição de dadores ao longo dos anos, os motivos que levam à inutilização das colheitas, quais os tipos de sangue mais comuns, as faixas etárias que mais se voluntariam, onde se concentram o maior número de dadores, etc. É o conjunto destes indicadores que permite identificar causas até então desconhecidas.

Foi desenvolvido um sistema de informação de BI de forma a solucionar estas necessidades e onde a informação acerca das Dádivas se encontra centralizada. Desta forma, o sistema encontra-se preparado para necessidades futuras sem ter de se fazer grandes alterações à estrutura do mesmo.

De forma a que os profissionais de saúde possam acompanhar os indicadores implementados, foi disponibilizado um conjunto de relatórios, nos quais a informação pode ser filtrada e se encontra sob a forma de dashboards. Os profissionais de saúde têm à sua disposição relatórios com os dados dos dadores, das colheitas e do resultado de análises de sangue.

Por fim, como forma de validar o sistema desenvolvido, foi elaborado inquéritos de satisfação a profissionais de IT e de saúde, onde se avalia a usabilidade e a utilidade do mesmo. Ambos os grupos demonstraram opiniões bastante positivas às afirmações que lhes foram colocadas acerca da aplicação desenvolvida. De acordo com os resultados obtidos, o desenvolvimento da aplicação foi conseguido segundo os objetivos a atingir e é de interesse utilizar este tipo de aplicação.

Palavras-chave: Dádivas de Sangue, dadores, colheitas, Análises de Sangue, Business Intelligence, indicadores, relatórios, dashboards

Abstract

The organization that provided the data to prepare this dissertation is responsible for the management and monitoring of Blood Donations at a national level. It stores data regarding donors, blood collections, and blood test results.

All of this information becomes useful for healthcare professionals as they are responsible for improvements related to healthcare. Thus, the way information is made available is of utmost importance, as it is of interest to them to obtain the data as quickly as possible to act accordingly.

A set of indicators on donors, collections and blood tests results have been developed to draw conclusions about the increase/decrease of donors over the years, the reasons leading to the disposal of collections, what types of blood are the most common, the most voluntary age groups, where the largest number of donors are concentrated, etc. It is the set of these indicators that makes it possible to identify hitherto unknown causes.

A BI information system has been developed to address these needs and where the information about the donations is centralized. This way, the system is prepared for future needs without having to make major changes to its structure.

In order to enable health professionals to follow the indicators implemented, a set of reports was made available, in which information can be filtered and is shown in the form of dashboards. Healthcare professionals have reports of donor data, collection data and blood test results available.

Finally, as a way to validate the developed system, satisfaction surveys were prepared for IT and health professionals, which assesses its usability and usefulness. Both groups expressed very positive opinions to the statements about the application developed. According to the results obtained, the application development was achieved according to the goals to be achieved and it is of interest to use this type of application.

Keywords: Blood Donations, donors, collections, Blood Tests, Business Intelligence, indicators, reports, dashboards

Agradecimentos

A realização desta dissertação contou com grande apoio e incentivos sem os quais não seria possível ter-se tornado realidade.

Em primeiro lugar, quero agradecer aos meus pais, pelo apoio incondicional, amizade e as opiniões prestadas ao longo desta caminhada. Um especial agradecimento à minha irmã que me incentivou e motivou para o Mestrado. Ao meu namorado, por todo o apoio, dedicação e por nunca me deixar desistir.

Agraço ainda ao meu Orientador, professor Paulo Oliveira, pelo apoio, disponibilidade, pelos ensinamentos e pelos conselhos ao longo desta dissertação.

Quero agradecer aos meus colegas de trabalho, por me terem apoiado e motivado nestes últimos dois anos.

E por último, agradeço também aos meus colegas de mestrado com os quais tive oportunidade de trabalhar e que partilharam comigo conhecimento.

“A maior recompensa para o trabalho do Homem não é o que ele ganha com isso, mas o que ele se torna com isso.”

John Ruskin

Índice

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Enquadramento..... | 1 |
| 1.2 | Problema | 2 |
| 1.3 | Motivação | 3 |
| 1.4 | Contribuições | 3 |
| 1.5 | Objetivos | 4 |
| 1.6 | Estrutura do Documento | 4 |
| 2 | Contexto | 7 |
| 2.1 | Descrição do problema | 7 |
| 2.2 | Análises de Sangue..... | 7 |
| 2.3 | Componentes Sanguíneos e Índices Presentes no Hemograma | 9 |
| 2.3.1 | Hemácias | 9 |
| 2.3.2 | Volume Globular Médio..... | 10 |
| 2.3.3 | Hematócrito | 10 |
| 2.3.4 | Hemoglobina | 11 |
| 2.3.5 | Hemoglobina Globular Média | 11 |
| 2.3.6 | Concentração de Hemoglobina Globular Média | 11 |
| 2.3.7 | Leucócitos..... | 12 |
| 2.3.8 | Plaquetas..... | 13 |
| 2.3.9 | Plasma | 13 |
| 2.4 | Tensão Arterial | 13 |
| 2.5 | Marcadores Serológicos..... | 14 |
| 2.5.1 | Hepatite B..... | 14 |
| 2.5.2 | Hepatite C..... | 15 |
| 2.5.3 | HIV | 16 |
| 2.5.4 | Sífilis | 16 |
| 2.5.5 | Alanina Aminotransferase..... | 17 |
| 3 | Estado de Arte | 19 |
| 3.1 | Business Intelligence..... | 19 |
| 3.1.1 | Processo de ETL | 20 |
| 3.1.2 | Armazém de Dados..... | 20 |
| 3.1.3 | Data Mart..... | 22 |
| 3.1.4 | Tabelas de Factos | 23 |
| 3.1.5 | Tabelas de Dimensão | 23 |
| 3.1.6 | Modelação Dimensional | 24 |
| 3.1.7 | Tipos de Modelos Dimensionais..... | 24 |
| 3.2 | OLAP | 25 |
| 3.2.1 | Sistema OLAP vs. Sistema OLTP..... | 26 |

| | | |
|----------|---|-----------|
| 3.3 | Ferramentas | 27 |
| 3.3.1 | Ferramentas de ETL..... | 28 |
| 3.3.2 | Ferramentas de análise de dados | 32 |
| 3.4 | BI na área da Saúde | 39 |
| 4 | Análise de valor | 41 |
| 4.1 | Inovar/Desenvolver Valor | 41 |
| 4.1.1 | Identificação da oportunidade | 42 |
| 4.1.2 | Análise da oportunidade | 42 |
| 4.1.3 | Geração de ideia e Enriquecimento | 43 |
| 4.1.4 | Seleção da ideia | 43 |
| 4.1.5 | Definição do Conceito | 43 |
| 4.2 | Análise de valor | 44 |
| 4.2.1 | Valor, valor para o cliente e valor percebido | 44 |
| 4.3 | Proposta de Valor | 44 |
| 4.4 | Modelo de Negócio CANVAS | 45 |
| 4.5 | Cadeia de valor de Porter..... | 46 |
| 4.6 | Analytic Hierarchy Process (AHP) | 48 |
| 4.6.1 | Divisão Hierárquica | 48 |
| 4.6.2 | Definição de Prioridades | 49 |
| 4.6.3 | Consistência Lógica | 49 |
| 5 | Design da Solução Proposta | 53 |
| 5.1 | Domínio das Dívidas de Sangue | 53 |
| 5.2 | Identificação dos Atores..... | 54 |
| 5.3 | Requisitos | 54 |
| 5.3.1 | Requisitos Não Funcionais..... | 54 |
| 5.3.2 | Requisitos Funcionais | 55 |
| 5.4 | Arquiteturas Possíveis | 57 |
| 5.5 | Arquitetura Selecionada..... | 58 |
| 5.5.1 | Fontes de dados | 59 |
| 5.5.2 | Staging Area..... | 60 |
| 5.5.3 | Modelação do DW | 62 |
| 6 | Implementação da solução | 68 |
| 6.1 | Tecnologias Utilizadas | 68 |
| 6.2 | Processo de ETL | 68 |
| 6.2.1 | Operações de Extração | 69 |
| 6.2.2 | Operações de Transformação..... | 70 |
| 6.2.3 | Operações de Carregamento | 75 |
| 6.3 | Análises de dados | 78 |
| 7 | Apresentação e Análise de Resultados | 79 |

| | | |
|-----------|---|------------|
| 7.1 | Indicadores..... | 79 |
| 7.2 | Dashboards..... | 79 |
| 7.2.1 | Informação Dadores..... | 79 |
| 7.2.2 | Informação Colheitas | 86 |
| 7.2.3 | Informação Resultado Análises de Sangue | 88 |
| 7.3 | Interface Mobile..... | 90 |
| 8 | Avaliação da Solução..... | 92 |
| 8.1 | Metodologia de Avaliação..... | 92 |
| 8.2 | Resultado dos Inquéritos realizados aos Profissionais de IT..... | 93 |
| 8.3 | Resultado dos Inquéritos realizados aos Profissionais de Saúde | 94 |
| 8.4 | Conclusão | 95 |
| 9 | Conclusão | 97 |
| 9.1 | Objetivos Alcançados | 97 |
| 9.2 | Problemas/Limitações..... | 98 |
| 9.3 | Trabalho Futuro | 98 |
| A1 | Anexo 1 - Logical Data Map (IPS - Staging Area) | 105 |
| A2 | Anexo 2 - Logical Data Map (Staging Area - Data Mart)..... | 111 |
| A3 | Anexo 3 - Control flow da solução desenvolvida | 117 |
| A4 | Anexo 4 - Visualização dos dados no Power BI..... | 119 |
| A5 | Anexo 5 - Interações Dashboards..... | 123 |
| A6 | Anexo 6 - Inquérito de satisfação aos profissionais de IT..... | 125 |
| A7 | Anexo 7 - Inquérito de satisfação aos profissionais de Saúde | 127 |

Lista de Figuras

| | |
|---|----|
| Figura 1 - Processo de ETL..... | 20 |
| Figura 2 - Abordagem Inmon | 21 |
| Figura 3 - Abordagem Kimball..... | 22 |
| Figura 4 - Tipos de Modelos Dimensionais | 25 |
| Figura 5 - Cubo OLAP (Microsoft, 2012)..... | 26 |
| Figura 6 - Quadrante Mágico das ferramentas ETL (Gartner, 2018b)..... | 28 |
| Figura 7 - Interface do Talend (Talend)..... | 29 |
| Figura 8 - Interface do SSIS (Michiels, Hans)..... | 30 |
| Figura 9 - Interface do ODI (Oracle) | 31 |
| Figura 10 - Quadrante Mágico de Business Intelligence (Gartner, 2018a). | 33 |
| Figura 11 - Interface do Sisense (Sisense)..... | 33 |
| Figura 12 - Interface do Qlik Sense (Qlik)..... | 34 |
| Figura 13 - Interface do Tableau (Tableau) | 35 |
| Figura 14 - Interface do PowerBI (Microsoft) | 35 |
| Figura 15 - Interface do Looker (Looker)..... | 36 |
| Figura 16 - Interface do IBM Cognos Analytics (IBM) | 37 |
| Figura 17 - Interface do Domo (Domo) | 37 |
| Figura 18 - Análise SWOT | 43 |
| Figura 19 - Modelo de Negócio CANVAS..... | 46 |
| Figura 20 - Cadeia de Valor de Porter (ResearchGate.net, 2017) | 46 |
| Figura 21 - Hierarquia AHP..... | 48 |
| Figura 22 - Diagrama de casos de uso..... | 55 |
| Figura 23 - Arquitetura 1..... | 58 |
| Figura 24 - Arquitetura 2..... | 58 |
| Figura 25 - Modelo de base de dados operacional..... | 59 |
| Figura 26 - Modelo da Staging Area | 61 |
| Figura 27 - Tabela de dimensão Data (DimDate) | 63 |
| Figura 28 - Tabela de dimensão Dadores..... | 64 |
| Figura 29 - Tabela de dimensão Colheitas | 65 |
| Figura 30 - Tabela de dimensão Junk | 66 |
| Figura 31 - Tabela de Factos Resultados Análises..... | 67 |
| Figura 32 – Representação do Data Mart | 67 |
| Figura 33 - Processo de ETL na arquitetura do sistema | 69 |
| Figura 34 - Fase de extração na arquitetura do sistema..... | 69 |
| Figura 35 - Fase de transformação na arquitetura do sistema | 70 |
| Figura 36 - Processo ETL da tabela "Dadores" | 71 |
| Figura 37 - Processo ETL da tabela "Colheitas" | 73 |
| Figura 38 - Processo ETL da tabela "Análises" | 74 |
| Figura 39 - Fase de carregamento na arquitetura do sistema..... | 75 |

| | |
|---|-----|
| Figura 40 - Carregamento da tabela DimDadores..... | 76 |
| Figura 41 - Carregamento da tabela de factos | 77 |
| Figura 42 - Fase de análise de dados na arquitetura do sistema | 78 |
| Figura 43 – Relatório informação dos dadores | 80 |
| Figura 44 - Drill down do Número de dadores por Ano | 81 |
| Figura 45 - Relatório informação dos dadores | 82 |
| Figura 46 – Relatório informação dos dadores | 84 |
| Figura 47 - Relatório de informação dos dadores | 85 |
| Figura 48 - Relatório de informação dos dadores | 86 |
| Figura 49 – Relatório informação das colheitas | 87 |
| Figura 50 – Relatório informação das colheitas | 88 |
| Figura 51 – Relatório informação das análises de sangue | 89 |
| Figura 52 – Relatório informação das análises de sangue | 89 |
| Figura 53 – Modo visualização mobile | 91 |
| Figura 54 - Seleção da fonte de dados no Power BI | 119 |
| Figura 55 - Representação do modelo relacional no Power BI | 120 |
| Figura 56 - Área de trabalho do Power BI | 121 |
| Figura 57 - Partilha de relatório no Power BI | 121 |
| Figura 58 - Interação em dashboard “Dadores” | 123 |
| Figura 59 - Interação dashboard "Colheitas" | 124 |
| Figura 60 - Interação dashboard "Resultados Análises de Sangue" | 124 |
| Figura 61 - Inquérito de Satisfação (Profissionais IT) | 125 |

Lista de Tabelas

| | |
|---|----|
| Tabela 1 - Compatibilidade Sanguínea..... | 8 |
| Tabela 2 - Níveis de Hemácias..... | 10 |
| Tabela 3 - Níveis de hematrócitos..... | 10 |
| Tabela 4 - Níveis de hemoglobina | 11 |
| Tabela 5 - Níveis de Leucócitos | 12 |
| Tabela 6 - Níveis de Neutrófilos | 12 |
| Tabela 7 - Níveis de plaquetas | 13 |
| Tabela 8 - Definições e classificações dos níveis de pressão arterial (SPHA, 2019)..... | 14 |
| Tabela 9 -Diferenças entre OLAP e OLTP (TechDifferences, 2016)..... | 27 |
| Tabela 10 - Comparação de ferramentas de ETL | 31 |
| Tabela 11 - Comparação entre ferramentas de BI..... | 38 |
| Tabela 12 - Tabela de Valores por Saaty (Santos, 2008)..... | 49 |
| Tabela 13 - Tabela com prioridades para critérios | 50 |
| Tabela 14 - Pesos dos Critérios | 50 |
| Tabela 15 - Critério Custo em relação às alternativas | 50 |
| Tabela 16 - Critério Confiabilidade em relação às alternativas | 50 |
| Tabela 17 - Critério Tempo Desenvolvimento em relação às alternativas | 51 |
| Tabela 18 - Pesos Finais das alternativas | 51 |
| Tabela 19 - Escala de Likert | 93 |
| Tabela 20 - Frequência de respostas dos profissionais de IT..... | 93 |
| Tabela 21 - Frequência das respostas dos profissionais de saúde..... | 94 |

Acrónimos e Símbolos

Lista de Acrónimos

| | |
|----------------|--|
| ETL | Extração, Transformação, Carregamento |
| BI | <i>Business Intelligence</i> |
| SNS | Serviço Nacional de Saúde |
| RBC | <i>Red Blood Cell Count</i> |
| MCV | <i>Mean Corpuscular Volume</i> |
| HCT | <i>Hematocrit</i> |
| HGB | <i>Hemoglobin</i> |
| MCH | <i>Mean Corpuscular Hemoglobin</i> |
| MCHC | <i>Mean Corpuscular Hemoglobin Concentration</i> |
| WBC | <i>White Blood Cell</i> |
| PLT | <i>Platelet</i> |
| U/L | <i>Units per Liter</i> |
| HIV/VIH | Vírus da Imunodeficiência Humana |
| ALT | <i>Alanina Aminotransferase</i> |
| DW | <i>Data Warehouse</i> |
| OLAP | <i>Online Analytical Processing</i> |
| OLTP | <i>Online Transactional Processing</i> |
| IT | <i>Information Technologies</i> |
| API | <i>Application Programming Interface</i> |
| SQL | <i>Structured Query Language</i> |
| PDF | <i>Portable Document Format</i> |
| ODS | <i>Operational Data Store</i> |

SCD *Slowly Changing Dimensions*

Lista de Símbolos

μ micro

1 Introdução

O presente documento resulta do desenvolvimento da dissertação de Mestrado de Engenharia Informática, no ramo de Sistemas de Informação e Conhecimento, do Instituto Superior da Engenharia do Porto.

É neste capítulo que, de forma sucinta, é efetuada a contextualização do tema, análise a Dádivas de Sangue de uma organização que procede à sua recolha e é apresentado o problema que esta enfrenta. Para além disso, é descrita a motivação que levou à escolha do tema e os objetivos que vão contribuir para a resolução do problema.

1.1 Enquadramento

A recolha de Dádivas de Sangue sempre foi um ato de grande importância na sociedade. Em Portugal, como forma de “homenagear os portugueses que efetuam as dádivas de sangue em prol dos doentes”, foi escolhido o dia 27 de março para se comemorar o dia Nacional do Dador (SNS, 2019).

São estes dadores, que permitem que as reservas de sangue se mantenham estáveis a nível nacional e possam ser utilizadas para salvar vidas. Existe um conjunto de direitos aos quais um dador de sangue pode ter acesso, realçando-se sobretudo a “salvaguarda da integridade física e mental, a confiabilidade dos dados e a isenção de taxas moderadoras no SNS” (IPST, 2017a).

Os sistemas de informação permitem que os dados sejam apresentados aos utilizadores de forma simples e útil. Na área da saúde, estes sistemas permitem que a informação fique armazenada de forma centralizada e que possa ser acedida pelas várias instituições de prestação de cuidados de Saúde.

A organização em estudo nesta dissertação possui informação relativa à atividade de recolha de sangue a nível nacional e responsabiliza-se por monitorizar toda essa informação. De certa forma, esta encarrega-se pela obtenção e gestão de informação acerca de dadores, colheitas, reservas e consumo de componentes sanguíneos.

A monitorização que a organização efetua, é realizada de forma contínua para que as demais instituições possam obter a informação sempre atualizada.

Ao longo de vários anos, informações sobre as Dádivas de Sangue foram adquiridas de diversas fontes que originaram dados inconsistentes devido ao formato das fontes ser diferente e de não seguirem um único padrão. Para além disto, a falta de exploração nos dados permitiu que muita informação essencial para a obtenção de conhecimento fosse descartada, impossibilitando novas descobertas.

Nesta dissertação vão ser analisados dados relativos às Dádivas de Sangue para se poder retirar conclusões. Na análise a efetuar é essencial que os dados estejam consistentes, se encontrem atualizados e que haja um histórico associado a cada utente, para que este possa ser devidamente acompanhado e detetadas atempadamente, caso existam, valores anómalos nos seus componentes sanguíneos.

Para além disto, e tendo em conta que o utilizador final são os profissionais de saúde, estes conseguem ter o quotidiano mais facilitado graças à simplificação e desmaterialização de processos até então usados. A forma como lhes é apresentada a informação dos utentes é um fator importante, pois o tempo de análise aos dados diminuiu e a tomada de decisão é conseguida mais rapidamente.

1.2 Problema

A organização de recolha de sangue a nível nacional, tem de ter a informação atualizada relativamente às dádivas.

Como forma de conseguir promover as Dádivas de Sangue existem aplicações que possibilitam aos utilizadores obter informações sobre a gestão das mesmas em tempo real, visualizar o estado das reservas a nível nacional e obter informações sobre local e horário para efetuar a dádiva de sangue. Para além disso, permitem a gestão da dádiva, pois é possível agendar o momento mais oportuno para tal e utilizam técnicas de ludificação, como forma de recompensa pelos utilizadores divulgarem a aplicação nas redes sociais.

Apesar de existirem este tipo de aplicações, não existe algo que permita visualizar dados estatísticos automaticamente sobre dádivas de sangue efetuadas ao longo do tempo.

Perceber se o número de dádivas tem vindo a aumentar/diminuir, se existem dádivas rejeitadas e quais os motivos, perceber o tipo de sangue mais comum de entre os doadores e se os valores dos marcadores Serológicos se encontram dentro dos parâmetros normais estabelecidos pela Direção Geral de Saúde, perceber se houve aumentos de dádivas em determinadas regiões e quais as faixas etárias que mais se voluntariam são alguns dos exemplos de indicadores que deveriam estar disponíveis em tempo real.

Para além disso, conseguir que a informação esteja desta forma organizada torna-se vantajoso para a organização na medida em que quando for necessário efetuar auditorias internas ou externas, o processo é simples porque a estrutura de armazenamento de dados possibilita obter facilmente a informação.

1.3 Motivação

O assunto apresentado constitui um problema recorrente no Sistema Nacional de Saúde que tem originado constrangimentos aos utentes e aos profissionais de saúde.

A escolha do tema para a dissertação deve-se ao facto de existir da parte da autora deste trabalho um gosto enorme pela área da saúde e vontade de contribuir com melhorias nos processos atuais. Para além disto, acresce a experiência obtida nos últimos três anos em projetos do Sistema Nacional de Saúde e o gosto pelo tema de qualidade de dados, que ultimamente tem sido tratado com grande importância nas organizações. A qualidade dos dados corresponde ao valor atribuído à informação disponibilizada pelas organizações, pelo que estas devem mantê-la sempre atualizada e válida de forma a contribuir para o bom funcionamento da empresa quer a nível dos processos transacionais, operacionais, estratégias de marketing e negócios” (GS1, 2017).

A solução a desenvolver torna-se atrativa porque influencia nas melhorias nos Cuidados de saúde. Para além de elaborar a arquitetura sobre este sistema, existe a necessidade de proceder à sua implementação, utilizando ferramentas Microsoft para a integração, tratamento e análise de dados.

1.4 Contribuições

O sistema a desenvolver baseia-se num sistema de Business Intelligence e tem o intuito de tornar a tomada de decisão dos profissionais de saúde mais eficiente.

O sistema permite realizar a gestão acerca de Dádivas de Sangue, nomeadamente possibilita a gestão de dados administrativos (informação pessoal dos dadores, dados demográficos, instituições) e dados clínicos (dados relativos a análises de sangue efetuadas aos dadores).

Este tipo de gestão sobre os dados contribui para otimizar o processo de análises dos dados da instituição em causa, devido à possibilidade de se poder obter relatórios facilmente sobre qualquer assunto, bem como visualizar de imediato os resultados através de dashboards.

São as dashboards que permitem visualizar e acompanhar em tempo real todas as atividades efetuadas pela organização através de gráficos e de tabelas interativas, enquanto que os relatórios são apenas o resultado de uma ação num período de tempo, sendo sempre necessário gerar novos quando se verificam alterações nos dados ou caso o período de tempo seja modificado (Keep.I, 2018).

É através deste acompanhamento, que se consegue perceber as necessidades atuais e futuras de acordo com o histórico de dados e atuar mediante essa informação.

1.5 Objetivos

Esta dissertação tem como principal objetivo desenvolver um sistema de informação de Business Intelligence sobre dados de Dádivas de Sangue de uma instituição, de forma a melhorar a qualidade da informação e obter conhecimento nos dados.

O sistema deve refletir a informação sob forma de relatórios e dashboards, para facilitar a leitura dos dados aos profissionais de saúde, para que estes consigam retirar de imediato conclusões acerca de determinados indicadores de Dádivas de sangue, apoiar nas decisões e auxiliar na deteção/correção de anomalias. Também para os utentes a implementação deste sistema se torna vantajoso indiretamente, pois como o tempo de resposta na tomada de decisão é menor, produzem melhorias a nível de diagnóstico e de prevenção.

Posto isto, e de acordo com o problema pretende-se:

- Desenvolver um Data Mart que contenha informações acerca das Dádivas de sangue.
- Construir o processo de ETL - Extração, Transformação e Carregamentos de dados.
- Efetuar análises aos dados.
- Apresentar os dados de forma interativa ao utilizador final.

1.6 Estrutura do Documento

Este documento encontra-se estruturado em nove capítulos.

No primeiro capítulo é apresentado e efetuada a contextualização do problema. Para além disso, são apresentadas as motivações que levaram à escolha do tema, as contribuições para a resolução do problema, os objetivos a atingir e a estrutura do documento.

No segundo capítulo, é feito um breve estudo acerca de componentes sanguíneos e dos valores padrão estabelecidos pela Direção Geral de Saúde para índices presentes no Hemograma e marcadores serológicos que permitem fazer um diagnóstico acerca do estado de saúde do dador. Este capítulo é essencial para perceber os conceitos que permitem retirar conclusões através dos resultados obtidos ao longo desta dissertação.

No terceiro capítulo, é apresentado o estado da arte, ou seja, é onde é efetuado uma revisão da literatura sobre conceitos de Business Intelligence, a descrição do processo de ETL e apresentado diferentes arquiteturas de Data Warehouse. Para além disso, são apresentadas diferentes ferramentas de integração e de análise de dados que podem ser aplicadas ao sistema a desenvolver, sendo para tal realizadas comparações entre as mesmas.

No quarto capítulo, é efetuada a análise de valor ao tema proposto, através de utilização de ferramentas que existem no mercado e que procedem a essa avaliação.

No quinto capítulo, é efetuado o levantamento de requisitos (funcionais e não funcionais) que são necessários para desenvolver o sistema. É ainda neste capítulo que é apresentada uma

arquitetura para o sistema a desenvolver, os atores que vão interagir com o sistema e ainda, de acordo com o tema, são descritos os componentes que constituem a arquitetura do sistema.

No sexto capítulo são apresentadas as tecnologias que foram utilizadas na implementação do sistema, bem como todas as operações necessárias no processo de ETL ao longo da implementação do sistema, de acordo com a arquitetura que foi proposta desenvolver.

No sétimo capítulo são apresentados os resultados obtidos através de dashboards e são discutidos os valores apresentados. É ainda mostrado como se pode obter as dashboards na interface mobile com a ferramenta utilizada para a análise de dados.

No oitavo capítulo avalia-se a solução, através de questionários de satisfação efetuados a profissionais de IT e profissionais de saúde.

No nono capítulo são apresentadas as conclusões que se obtiveram com esta dissertação , nomeadamente os objetivos atingidos, os problemas/limitações encontradas e ainda é exposto o trabalho futuro que se poderia desenvolver, de acordo com o tema em estudo.

2 Contexto

O presente capítulo tem por objetivo, contextualizar melhor o problema, o qual será alvo de estudo ao longo desta dissertação e apresentar conceitos relacionados com componentes sanguíneos que se encontram presentes em análises de sangue. É essencial perceber estes conceitos e os valores entre os quais os componentes se consideram normais/anormais para se entender os resultados finais e retirar conclusões.

2.1 Descrição do problema

A organização apesar de efetuar estudos acerca das dádivas de sangue, não torna público alguns indicadores que são importantes para acompanhar ao longo do tempo o estado de saúde dos utentes a nível nacional.

De certa forma, existe uma enorme quantidade de dados que são recolhidos, no entanto como não existe um devido tratamento e análise aos mesmos, existe um conjunto de informações que se tornam inúteis para a obtenção de conhecimento.

Face ao exposto, é proposto um Sistema de Business Intelligence (BI) para explorar dados acerca de dádivas de sangues oriundos desta entidade, para se poder tirar conclusões que são importantes para o Serviço Nacional de Saúde.

Foi sobretudo a pensar nas dificuldades que os profissionais de saúde enfrentam devido às quantidades de informação dispersas por diferentes fontes de dados que se idealizou este tipo de sistema.

O sistema desenvolvido é centralizado e permite obter a informação sob a forma de relatórios e dashboards para facilitar o quotidiano dos profissionais de saúde. Assim sendo, a informação encontra-se sempre disponível e de fácil acesso, permitindo assim obter mais tempo para se dedicarem à tomada de decisão.

2.2 Análises de Sangue

O sangue é uma mistura de várias células suspensas num líquido chamado plasma que tem por objetivo defender o organismo de agentes nocivos. Para tal tem de efetuar o transporte de oxigénio e nutrientes para as células, remover substâncias resultantes da atividade celular e conduzir hormonas pelo organismo (Magalhães, 2018).

Quando se efetua a Dádiva de Sangue, o sangue é analisado e sujeito a processos para separação dos seus componentes, para mais tarde poderem ser administrados a quem deles precisa.

Dos componentes que constituem o sangue, o plasma representa sensivelmente 55% do volume e caracteriza-se por ter uma cor amarelada (possui 91% de água, 7% de proteínas e 2%










de gases e nutrientes) e os restantes 45% são representados por Células sanguíneas (Plaquetas, Glóbulos Brancos, Glóbulos Vermelhos) (Unidos pela Hemofilia, 2019).

Em 2016, Portugal registou 334.022 dádivas de sangue, o que corresponde a 15.300 litros de sangue doados por 256.073 dadores. Do total de dadores destacaram-se 30.660 por efetuarem a sua dádiva pela primeira vez (Leiderfarb, 2018).

Existe um sistema designado de ABO que se responsabiliza por classificar o sangue de acordo com quatros tipos de sangue: A, B, AB e O e dois subgrupos: RH Positivo e RH negativo (Lemos, 2018e).

A Tabela 1 mostra as possíveis relações entre tipos de sangue:

Tabela 1 - Compatibilidade Sanguínea

| Compatibilidade de Sangue | | DOADOR | | | |
|---------------------------|----|---|---|--|---|
| | | O | B | A | AB |
| RECECTOR | AB |  |  |  |  |
| | A |  | |  | |
| | B |  |  | | |
| | O |  | | | |

De acordo com os sistemas ABO e mediante a representação efetuada na tabela 1, pode-se entender o seguinte relativamente aos tipos de sangue:

Sangue A – Indivíduos que possuem este tipo de sangue podem receber de A e de O e podem doar sangue para A e AB

Sangue B – Indivíduos com este tipo de sangue podem receber de B e de O e podem doar para B e AB

Sangue AB - Indivíduos com este tipo de sangue podem receber de A, B, AB, O e podem doar para AB

Sangue O - Indivíduos com este tipo de sangue podem receber de O e doar para A, B, AB e O

De acordo com o Serviço Nacional de Saúde, para se ser dador de sangue é necessário que a pessoa possua bons hábitos de saúde, hábitos de vida saudáveis, peso igual ou superior a 50 kg e idade entre os 18 e os 65 anos. Para quem efetua a dádiva pela primeira vez, a idade limite é de 60 anos. Para além disso, por ano os homens podem doar sangue quatro vezes (de três em três meses) e as mulheres podem doar três vezes (de quatro em quatro meses) (SNS, 2016) .

Existem diferentes tipos de dádivas: Aférese, Autóloga e Homóloga (Diário da República, 2007).

A primeira, **Dádiva Aférese**, corresponde à recolha de componentes sanguíneos específicos do sangue total. Este tipo de dádiva tenta combater o “desperdício” e é aquela em que o processo costuma ser mais longo.

A **Dádiva Autóloga** refere-se à colheita que um determinado dador efetua para si próprio. Este tipo de dádiva requer autorização médica e tem de haver concordância pelo utente. Nesse sentido, o utente pode ser tratado com o próprio plasma ou componentes caso autorize a transformação do mesmo em derivados.

Por fim, a **Dádiva Homóloga**, requer que o sangue e os seus componentes sejam colhidos de um indivíduo para serem administrados em outro. Para além disso, podem sere usados em dispositivos médicos ou matéria-prima para fabrico de medicamentos.

2.3 Componentes Sanguíneos e Índices Presentes no Hemograma

Os componentes Sanguíneos são elementos celulares que se encontram no sangue, e que possuem funções de coagulação, de defesa contra infeções e de transporte de oxigénio e de dióxido de carbono. Estes componentes podem ser obtidos através de processos de centrifugação no sangue (Hemominas, 2014) .

Quando se realizam análises ao sangue, é frequentemente solicitado o Hemograma, para avaliar a morfologia e a constituição dos diferentes componentes sanguíneos, nomeadamente as Hemácias, os Leucócitos e as Plaquetas. Para além disso, avalia ainda a hemoglobina e o hematócrito (Deco Proteste, 2012).

Nas subsecções seguintes são apresentados os componentes Sanguíneos e os índices que estudam os mesmos e que se encontram presentes no Hemograma, de forma a perceber quais os valores de referência para cada um deles.

2.3.1 Hemácias

Este componente designa-se também de Glóbulos vermelhos ou eritrócitos e é o elemento que se encontra em maiores quantidades no sangue. Na sua constituição possui hemoglobina, glicose, água e enzimas. A cor vermelha do sangue deve-se à presença de grandes quantidades de hemoglobina, sendo esta a proteína que confere a cor vermelha ao mesmo. As Hemácias desempenham no organismo funções de transporte de oxigénio para as células e de dióxido de carbono para ser expelido.

São produzidas na medula óssea e possuem um tempo de duração no organismo de 120 dias (Significados, 2018).

Os valores de referência para os níveis normais de Hemácias no sangue (RBC), por milímetro cúbico, são apresentados na Tabela 2

Tabela 2 - Níveis de Hemácias

| Género | Níveis Normais de Hemácias (milhões/mm ³) |
|----------|---|
| Homens | 5 |
| Mulheres | 4,5 |

2.3.2 Volume Globular Médio

O Volume Globular Médio (MCV) é um índice presente no hemograma que indica a média do volume das Hemácias. Os valores de MCV são importantes para diagnosticar os vários tipos de anemia. Valores altos para MCV representam anemias por carência de ácido fólico (Vitamina B9), enquanto valores baixos para MCV representam anemias por carência de ferro (vitamina B12) (BioCampello, 2019).

O ácido Fólico e o Ferro são vitaminas do complexo B que se encontram presentes nas frutas, Leguminosas, Carne (Stuppiello, 2018). Considera-se que os valores normais para o volume corpuscular médio se encontram compreendidos entre 80 a 100 fentolitros, sendo os restantes valores considerados anormais (Lemos, 2018a).

2.3.3 Hematócrito

O Hematócrito (HCT) é um parâmetro que indica a percentagem de Hemácias no volume total do sangue (Lemos, 2019).

Nos casos em que os valores para o Hematócrito são baixos, pode ser indicativo de anemia, desnutrição, falta de vitamina B12 ou ácido fólico, leucemia ou excesso de hidratação. Pelo contrário, valores de Hematócrito altos, podem ser indicativos de desidratação, doenças pulmonares, doenças cardíacas.

A Tabela 3 representa os valores que se consideram normais e anormais para o Hematócrito.

Tabela 3 - Níveis de hematrócitos

| Género | Níveis normais de hematrócitos (%) |
|----------|--|
| Homens | Entre 40% e 50% |
| Mulheres | Entre 35% e 45% Em grávidas entre 34% e 47% |

2.3.4 Hemoglobina

A hemoglobina (HGB) é uma molécula que se encontra nas Hemácias e que confere a função de transporte de oxigénio.

Valores altos de Hemoglobina podem ser causados por desidratação, uso de tabaco, tumor nos rins, enquanto que valores baixos de Hemoglobina podem acontecer em caso de anemia, leucemia, insuficiência renal, hemorragia. Para além disto, os baixos valores podem ainda estar relacionados com a deficiência de ferro e vitaminas, uso de medicamentos para tratar cancro (Lemos, 2018b).

Os valores de referência da hemoglobina, encontram-se na Tabela 4.

Tabela 4 - Níveis de hemoglobina

| Género | Níveis normais de hemoglobina (g/dl) |
|----------|--------------------------------------|
| Homens | 14 a 18 g/dl |
| Mulheres | 12 a 16 g/dl |

2.3.5 Hemoglobina Globular Média

O Índice para a Hemoglobina Globular Média (MCH) indica a quantidade média de hemoglobina presente nas Hemácias (Rosa, 2018).

Quando MCH apresenta valores baixos deve-se à presença de anemia por falta de hemoglobina, sendo a mais habitual a anemia por défice de ferro.

Os valores altos para MCH correspondem à presença de anemia hipercrómica e ao alto défice de ácido fólico e vitamina B12. Apesar disto, os valores de referência para MCH encontram-se compreendidos entre 27 a 33 pc (picogramas).

2.3.6 Concentração de Hemoglobina Globular Média

A Concentração de Hemoglobina Globular Média (MCHC) avalia a concentração de hemoglobina dentro da Hemácia (BioCampello, 2019).

Quando os valores para MCHC se encontram acima dos valores normais, significa anemia hipercrómica, alterações da tiroide ou alcoolismo. Nestas condições, as hemácias são maiores, levando a anemia causada pela falta de vitamina B12 e ácido fólico.

Quando os valores se encontram abaixo do normal, indica anemia hipocrómica, que se deve à falta de ferro. Nesta situação, as Hemácias são menores que o normal, dado o valor médio de hemoglobina ser baixo.

Os valores que se consideram de referência para a concentração de hemoglobina globular médica encontram-se entre 32 a 36g de hemoglobina/dl (Dextro, 2019).

2.3.7 Leucócitos

Os leucócitos são também designados por glóbulos brancos. São produzidos na medula óssea e nos tecidos linfáticos e têm como função defender o organismo de agentes infecciosos, como por exemplo vírus, bactérias ou substâncias que causam alergias.

A Tabela 5 representa os valores de referência para os níveis normais de Leucócitos no sangue (WBC), por milímetro cúbico (Lemos, 2018a).

Tabela 5 - Níveis de Leucócitos

| Nível (mm ³) | Classificação |
|--------------------------|---|
| Menor que 4500 | Anemia, uso de antibióticos e diuréticos, má nutrição ou sistema imune fraco provocado por HIV, leucemia, lúpus ou quimioterapia. |
| Entre 4500 e 110000 | Normal |
| Maior que 11000 | Infeção ou doença recente, excesso de stress, efeito colateral de um remédio, alergias, artrite reumatoide, mielofibrose ou leucemia. |

2.3.7.1 Neutrófilos

Os neutrófilos (NEU) são um tipo de leucócitos e como tal são responsáveis pela defesa e imunidade do organismo. Têm como função, eliminar células doentes e combatem bactérias e fungos causadores de infeções (Reis, 2019).

Os neutrófilos são armazenados e produzidos na medula óssea e seguem os seguintes valores de referência:

Tabela 6 - Níveis de Neutrófilos

| Nível (mm ³) | Classificação |
|--------------------------|---|
| Menor que 1600 | Neutropenia |
| Entre 1600 e 8000 | Normal |
| Maior que 11000 | Neutrófila (devido ao stress, uso de medicamentos com adrenalina, exercício físico exagerado) |

2.3.8 Plaquetas

As plaquetas (PLT) são também designadas de trombócitos e são produzidas na medula óssea. São responsáveis pelo processo de coagulação do sangue, nomeadamente são produzidas em maiores quantidades quando existe perda de sangue excessiva.

O baixo valor de plaquetas encontra-se relacionado com infeções, uso de medicamentos, doenças relacionadas com imunidade.

O alto valor de plaquetas relaciona-se com altitude elevada, trabalho de parto, stress, exercício em excesso (Lemos, 2018c).

Tabela 7 - Níveis de plaquetas

| Nível (plaquetas/ μ L) | Classificação |
|----------------------------|----------------------------|
| Menor que 20 000 | Trombocitopenia |
| Entre 150 000 e 450 000 | Normal |
| Maior que 450 000 | Plaquetose ou trombocitose |

2.3.9 Plasma

O plasma é constituído por água (95%), sais minerais, proteínas, hormonas, nutrientes, gases e excreções (5%). Tem como funções o transporte das substâncias que o integram pelo organismo.

Possui como principais proteínas a albumina, que regula a pressão osmótica no sangue e a hemoglobina que fornece anticorpos ao organismo (UOL, 2018).

2.4 Tensão Arterial

Designa-se Tensão Arterial como sendo “pressão a que o sangue circula nas artérias”.

Os seus valores variam ao longo do dia e quando são normais, é assegurado que o sangue circula por todo o corpo, no entanto os valores podem encontrar-se abaixo/acima relativamente a esses valores.

Existem dois parâmetros sobre os quais assenta a medição da pressão arterial:

- Pressão Arterial Mínima (“diastólica”)
- Pressão Arterial Máxima (“Sistólica”)

A Pressão Arterial Mínima é a pressão que o sangue exerce nas artérias quando o coração está relaxado e a pressão Arterial Máxima corresponde à pressão que o sangue exerce nas paredes

das artérias quando o coração está a bombear sangue. A pressão arterial ideal deve ser inferior ou igual a 120/80, em que o primeiro número diz respeito à pressão arterial máxima e o segundo à pressão arterial mínima. Acima destes valores existe risco de doenças coronária ou acidente vascular cerebral. Esta expressa-se em milímetros Cúbicos (mmHg).

A Tabela 8 apresenta os valores para a pressão arterial.

Tabela 8 - Definições e classificações dos níveis de pressão arterial (SPHA, 2019)

| CATEGORIA | SISTÓLICA | | DIASTÓLICA |
|-------------------------------|-----------|------|------------|
| Ótima | < 120 | e | < 80 |
| Normal | 120-129 | e/ou | 80-84 |
| Normal alta | 130-139 | e/ou | 85-89 |
| Hipertensão grau 1 | 140-159 | e/ou | 90-99 |
| Hipertensão grau 2 | 160-179 | e/ou | 100-109 |
| Hipertensão grau 3 | ≥ 180 | e/ou | ≥ 110 |
| Hipertensão sistólica isolada | ≥ 140 | e | < 90 |

É importante saber os valores para a tensão arterial, antes de se realizar a dádiva de sangue, uma vez que esta influência a mesma. Caso o individuo seja hipotenso a dádiva não pode ser realizada e quando se está perante um individuo hipertenso a dádiva pode acontecer caso os valores da tensão arterial estejam controlados.

No momento da dádiva, é recomendado que os valores da tensão arterial estejam compreendidos entre (IPST, 2018):

Sistólica - igual ou superior a 100 mmHg e igual ou inferior a 180 mmHg

Diastólica - igual ou superior a 60 mmHg e igual ou inferior a 100 mmHg

2.5 Marcadores Serológicos

É através de marcadores serológicos que se consegue obter o diagnóstico acerca de infeções provocadas pelas Hepatites Virais. Estes identificam o determinante antigénico envolvido, distinguem em que fase a infeção se encontra (fase aguda ou fase crónica), avaliam a infetividade e fazem o prognóstico. Nas Subsecções seguintes, são apresentados algumas Hepatites virais, que constituem um problema de saúde pública.

2.5.1 Hepatite B

Existe um conjunto de marcadores serológicos que são utilizados para o diagnóstico da Hepatite B, para ajudar no acompanhamento dos pacientes e na monitorização de tratamento (Câmara, 2015).

- **HBsAg** (Antígeno de Superfície do Vírus B)
A presença deste indicador indica infeção, podendo ser esta aguda ou crónica. Aparece geralmente 1 a 2 semanas após o indivíduo se expor ao vírus.
- **Anti-HBc IgM** (Anticorpo IgM contra o antígeno core do Vírus B)
A presença deste indica infeção aguda e pode ser encontrado no soro até 32 semanas após infeção.
- **Anti-HBc total** (Anticorpos totais contra o antígeno core do Vírus B)
Este marcador encontra-se presente na infeção aguda, na infeção crónica e em pacientes que já se recuperaram da doença (indica contacto prévio com o vírus).
- **HBeAg** (Antígeno “e” do Vírus B)
Este marcador, quando positivo indica infeção. Na infeção aguda, desaparece entre 2 a 6 semanas. Caso persista após 8 a 10 semanas, existe uma forte possibilidade de evoluir para infeção crónica.
- **Anti-HBe** (Anticorpo contra o antígeno “e” do vírus B)
Este indicador é detetado geralmente após o desaparecimento do HbeAg. Sugere redução ou ausência de vírus. O Anti-HBe consegue-se detetar por 1 a 2 anos, no entanto em alguns indivíduos não é detetável.
- **Anti-HBs** (Anticorpo contra o antígeno de superfície do vírus B)
Encontra-se presente geralmente após o desaparecimento do HbsAg e indica resolução da infeção e imunidade. Em pessoas vacinadas pode-se encontrar este indicador de forma isolada.

2.5.2 Hepatite C

A Hepatite C é uma doença viral que origina infeções no fígado e raramente manifesta sintomas. Esta doença transmite-se através do contacto com sangue contaminado, por transfusões, material e drogas injetáveis.

Existem dois indicadores de Hepatite C:

- **Anti-HCV**
Este indicador permite saber se o resultado de Hepatite C é positivo ou negativo. No caso de ser positivo, são necessários outros exames para que se detete o material genético do vírus no sangue. Quando o resultado dos segundos exames solicitados é negativo, indica que o indivíduo já esteve em contacto com o vírus, só que o mesmo se curou.

Caso o Anti-HCV for negativo, o indivíduo em causa não teve contacto com o vírus (Geovana, 2015).

- **HCV-RNA**

Este marcador é o primeiro a surgir, manifestando-se entre uma ou duas semanas após infeção. Efetua-se o teste para detetar, quantificar e caracterizar o material genético do vírus, após se obter resultado positivo através do Anti-HCV.

2.5.3 HIV

O Vírus de Imunodeficiência Humana (VIH) é o vírus associado ao Síndrome de Imunodeficiência Adquirida (SIDA) que ataca progressivamente o sistema imunitário fazendo com que a pessoa se torne mais vulnerável a infeções e cancro (HIV.gov, 2017b).

O HIV geralmente é transmitido através de relações sexuais, troca de seringas e amamentação (HIV.gov, 2017a).

Existe um indicador sorológico para o HIV:

- **Anti-HIV**

A presença deste anticorpo no sangue permite verificar se o indivíduo é portador do vírus. Estes anticorpos geralmente só começam a ser produzidos cerca de 3 meses após o contacto com o vírus.

2.5.4 Sífilis

Sífilis é uma infeção que se transmite sexualmente e é detetada quando se efetuam exames para diagnosticar a presença do vírus.

VDRL (Venereal Disease Research Laboratory) é o exame mais utilizado para diagnosticar a Sífilis e é utilizado como método de rastreio. O resultado é um indicador semi-quantitativo, pois representa-se sobre formas de diluição (Ex: 1/8 significa que ainda são detetados anticorpos após se diluir o sangue 8 vezes).

Após a infeção, o resultado permanece positivo entre 4 a 6 semanas e em alguns casos, quando na presença de outras doenças, pode-se obter casos de falsos positivos ou falsos negativos caso o exame seja feito entre 1 a 2 dias após a infeção.

FTA_ABS (Fluorescent Treponemal Antibody Absorption) ou **TPHA** (Treponema pallidum hemagglutination assay) é um exame mais específico que o VDRL e considera-se um exame de confirmação, pois apresenta uma menor taxa para falsos positivos comparativamente com o VDRL.

No caso de ser negativo, significa que não existe presença de anticorpos e, portanto, o indivíduo em causa não se encontra infetado e /ou não esteve em contacto com a doença.

No caso de ser positiva, considera-se que o indivíduo quer no passado ou no presente já esteve em contacto com o Vírus.

Em suma, segue abaixo um resumo com a interpretação de resultados dos dois exames:

VDRL e FTA-ABS. Positivos – Presença de Sífilis

VDRL Positivo e FTA-ABS Negativo – Existência de outra doença que não Sífilis

VDRL Negativo e FTA-ABS Positivo- Sífilis na fase inicial ou já ultrapassada

VDRL e FTA-ABS negativos – o indivíduo não possui Sífilis

2.5.5 Alanina Aminotransferase

Quando se pretende detetar lesões e doenças no fígado é utilizado um exame para medir as quantidades de Alanina Aminotransferase (ALT) no sangue.

É através da elevada presença desta enzima no sangue que se consegue detetar esses problemas, pois quando existem lesões provocadas por vírus ou até mesmo substâncias tóxicas esta enzima é libertada para a corrente sanguínea (Lemos, 2018d).

Geralmente encontra-se entre 7 e 56 U/L de sangue, no entanto existem valores que são considerados anormais:

ALT muito alta

- 10 vezes superior ao valor normal - pode ter origem em uma hepatite aguda causada pelo vírus ou por alguns medicamentos
- 100 vezes superior à normal – é comum em indivíduos que consomem substâncias que causam danos no fígado, nomeadamente drogas e álcool.

ALT elevada

- 4 vezes superior ao normal - Pode indicar hepatite crônica como cirrose ou cancro.

3 Estado de Arte

Este capítulo pretende descrever os conceitos subjacentes a sistemas de Business Intelligence (BI) e explicar como estes apoiam a tomada de decisão nas organizações.

Ao longo deste capítulo são apresentadas as etapas do processo de ETL, os componentes do sistema de BI, arquiteturas, tecnologias, ferramentas e efetuada a comparação entre diferentes softwares.

Por fim, é mostrado como os sistemas de BI constituem uma mais valia para a área da saúde.

3.1 Business Intelligence

O termo Business Intelligence teve a sua origem em 1989 e o mesmo foi promovido por Howard Dresner quando era analista no Instituto Gartner (Power, 2007).

O conceito de Business Intelligence (BI), tem-se utilizado durante vários anos, no entanto não existe uma definição concreta que seja aceite na sua totalidade pela indústria, já que o mesmo é composto por um conjunto de processos (C-Studio, 2017).

Mesmo assim, BI pode definir-se como sendo práticas de gestão implementadas com recurso a software com o intuito de apoiar e rentabilizar a tomada de decisão, estabelecer os objetivos do negócio e aumentar a liderança das organizações (Lapa, Bernardino and Figueiredo, 2014).

De certa forma, os sistemas de Business Intelligence são sistemas de suporte à decisão (DSS) orientados por dados, que utilizam tecnologia para poder analisá-los. É através dos resultados obtidos que se consegue obter informações úteis para o negócio e apoiar na tomada de decisão (Teixeira, Filho and Costa, 2018). As grandes quantidades de informação sofrem um processo de transformação, através de ferramentas ETL (Extract, Transform, Load) para mais tarde os dados possam ser representados visualmente, de forma apelativa, transmitindo os pontos essenciais sobre o negócio, nomeadamente tendências e problemas que eram até então desconhecidos. Isto possibilita às organizações medirem as suas ações perante os resultados e desenvolver estratégias competitivas essenciais para se manterem no mercado.

Os sistemas de Business Intelligence permitem aumentar a produtividade nas organizações, através de processos automatizados para conseguir obter relatórios, tabelas e gráficos acerca do negócio. Isto permite reduzir o número de recursos e o tempo de análise. Qualquer pessoa, que não seja técnica consegue identificar rapidamente qual a área do negócio que necessita de melhorias. Apesar disto tudo, é de considerar que desenvolver um Sistema de BI requer custos, complexidade e tempo de implementação, o que pode ser entrave para muitas empresas, especialmente de pequeno e médio porte.

3.1.1 Processo de ETL

O processo de ETL é constituído por três etapas. Permite extrair, transformar e carregar dados de diferentes fontes para um repositório central, o qual irá ter a informação estruturada para mais tarde se fazer uso da mesma (Talend, 2019).

É através da etapa de extração, que os dados são retirados do sistema de origem, geralmente base de dados ou ficheiros internos ou externos à organização e destinam-se a uma área de preparação que se designa de “Staging Area”, onde são transformados.

A etapa de transformação permite converter os dados extraídos em dados que atendem os requisitos do sistema destino. Esta é a fase mais importante no processo de ETL e é onde se garante a qualidade dos dados. Para tal, existe um conjunto de tarefas a serem executadas como remover duplicados, garantir o formato dos dados, corrigir erros, eliminar dados que não são relevantes. Sem esta etapa, as organizações não podem confiar nos dados migrados.

A etapa de carregamento, permite concluir o processo de ETL. Os dados transformados são carregados no sistema destino (Data Warehouse e/ou Data Mart). A conclusão desta etapa depende do volume de dados, da estrutura dos mesmos e da frequência com que são carregados (Zhao, 2017).

A Figura 1 representa o processo de ETL.

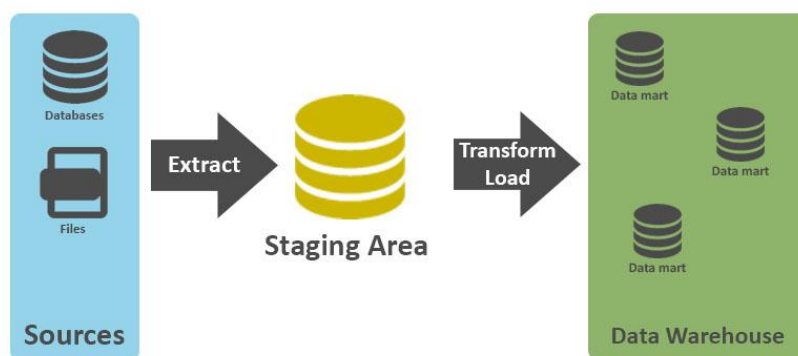


Figura 1 - Processo de ETL

3.1.2 Armazém de Dados

Armazém de Dados (Data Warehouse) é o núcleo do sistema de BI. É um repositório central que possui dados históricos e atuais relativos à organização e que já passaram pelo processo de ETL.

Nele, os dados já apresentam a devida qualidade e são organizados, para que se possa retirar conclusões acerca de determinada área de Negócio. Como a informação se concentra num só local, a performance do DW é maior comparativamente a sistemas operacionais.

Existem duas abordagens com as quais as organizações se deparam quando projetam o Data Warehouse. Uma das abordagens foi desenvolvida por Bill Inmon e a outra por Ralph Kimball. Ambos defendem que o DW é o repositório central que é carregado utilizando o processo de ETL. As diferenças residem na forma como os dados são modelados, carregados e armazenados no DW (Rangarajan, 2016).

3.1.2.1 Abordagem Inmon

Bill Inmon defende uma abordagem top-down (de cima para baixo) (George, 2012). O data warehouse é o componente mais importante do sistema e centraliza todos os dados vindos de vários sistemas operacionais, pelo que possui a informação normalizada. Após a sua construção, surgem os Data Marts quando é necessário obter algum segmento do negócio. A integridade dos dados é feita no Data Warehouse, seguindo a informação para cada Data Mart.



Figura 2 - Abordagem Inmon

3.1.2.2 Abordagem Kimball

Ralph Kimball defende uma abordagem bottom-up (de baixo para cima) (George, 2012), pois considera que é necessário ter em conta os processos de negócio mais importantes da organização. Para tal, defende que se deve criar primeiro os Data Marts que possuem os dados mais relevantes sobre cada área do negócio e o Data Warehouse constitui a combinação desses Data Marts.

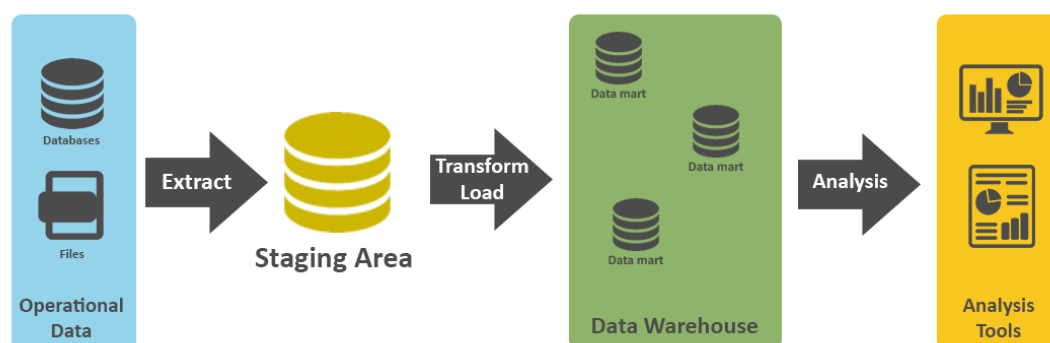


Figura 3 - Abordagem Kimball

Ambas as abordagens são válidas e podem ser aplicadas para projetar o Data Warehouse nas organizações, tendo em conta os objetivos traçados pelas mesmas. A abordagem de Inmon pode ser aplicada em organizações que dispõem de bastante tempo para a implementação do sistema de BI e que já se encontram estáveis economicamente, enquanto que a abordagem de Kimball pode ser aplicada em organizações que pretendem demorar menos tempo na implementação e cujo foco é a otimização (George, 2012).

3.1.3 Data Mart

Os Data Marts constituem subconjuntos de dados dos Data Warehouse. Estes são orientados por assunto, uma vez que se referem a um departamento específico da empresa.

Os repositórios obtêm os dados desnormalizados e indexados a partir do DW para conseguirem efetuar múltiplas consultas em simultâneo e facilitar o processo de pesquisa, tornando o mesmo mais rápido.

Estes repositórios podem ser construídos a partir de um data warehouse (caso a abordagem aplicada seja a de Inmon) ou podem ser contruídos através de sistemas operacionais internos ou externos (abordagem de Kimball).

É com base na relação que o Data Mart tem com os dados de origem que levam à existência de três tipos de Data Marts: Dependentes, Independentes e Híbridos.

Os Data Marts dependentes, são criados a partir de um Data Warehouse existente. Todos os dados do negócio são armazenados centralmente e só depois se extrai os dados que são necessários para as análises de determinada área de negócio da organização.

Os Data Marts Independentes, são criados através de dados internos e/ou externos, processados e carregados no Data Mart. Tornam-se benéficos quando se pretende atingir metas de curto prazo. Apesar disto, como cada um utiliza uma determinada lógica e ferramentas de ETL, tornam o processo mais complexo e mais difícil de gerir.

Por fim, os data marts Híbridos combinam dados do Data Warehouse e de sistemas operacionais. Este tipo de Data Mart foca-se na integração dos dados e na satisfação do utilizador final.

3.1.4 Tabelas de Factos

A tabela de factos é a principal tabela do Data Warehouse. Esta encontra-se ligada às dimensões e nela são armazenadas as métricas e as Foreign Keys. As métricas ou factos é tudo o que a organização consegue medir, quantificar e as Foreign Keys são as chaves estrangeiras que servem para relacionar os dados entre a tabela de factos e as tabelas de dimensão (Piton, 2017). Os factos que se encontram nesta tabela encontram-se definidos segundo um nível de granularidade, ou seja, são armazenados na tabela segundo um nível de detalhe definido.

3.1.5 Tabelas de Dimensão

São as tabelas de Dimensão que adicionam contexto aos factos. Estas encontram-se ligadas à tabela de factos através de uma chave denominada de chave substituta (Surrogate Key), que apresenta como vantagem o aumento da performance entre elas. Existem vários tipos de dimensões, das quais se destacam as dimensões denominadas Slowly Changing Dimensions e Junk Dimensions.

3.1.5.1 Slowly Changing Dimensions

Em 1996 foi introduzido por Kimball uma técnica denominada de Slowly Changing Dimensions (SCD), que permite classificar as atualizações dos atributos das dimensões de acordo com o tipo de mudança que se pretende que ocorra. Os tipos mais vulgares de SCD são: Tipo 1, Tipo 2 e Tipo 3 (NuWave Solutions, 2018).

O SCD de tipo 1 corresponde à alteração que não guarda nenhum histórico na dimensão. Trata-se do tipo mais simples, pois não existe controlo sobre a atualização de dados, apenas existe sobreposição de informação.

O SCD de tipo 2 é a técnica mais utilizada nas atualizações das dimensões. Com este tipo de SCD é adicionado um novo registo com as mudanças, preservando os dados anteriores.

Por último, o SCD de tipo 3 permite manter as alterações no mesmo registo. Esta técnica possibilita a adição de uma nova coluna com os dados novos, mantendo também na antiga coluna o valor anterior.

3.1.5.2 Junk Dimension

A dimensão Junk é muitas vezes associada ao próprio nome. Este tipo de dimensão é constituída por um conjunto de atributos (códigos transacionais, flags, atributos de texto) não relacionados, que não pertencem a nenhuma dimensão específica, nem à tabela de factos. Tem o objetivo de reduzir o número de dimensões, diminuir do espaço ocupado pela tabela de factos e ao mesmo tempo evita-se perda de informação.

3.1.6 Modelação Dimensional

Esta técnica é utilizada em data Warehouse e difere da modelação relacional. O conceito foi desenvolvido por Ralph Kimball e é composto por tabelas de “Factos” e de “Dimensão”.

O objetivo é conseguir obter a informação do Data Warehouse mais facilmente. Nos modelos relacionais, os dados são normalizados para que não haja redundância na informação, enquanto que os modelos dimensionais apenas organizam os dados para se obter mais facilmente a informação e conseguir transmitir a mesma ao utilizador final.

Na modelação dimensional, pode-se representar a informação como um cubo se tratasse, pois quanto mais nos aproximamos de cada dimensão, mais detalhe se consegue obter. As tabelas de dimensão ligam-se diretamente à tabela de factos e esta última encontra-se localizada centralmente.

3.1.7 Tipos de Modelos Dimensionais

Existem três tipos de modelos Dimensionais (Guru99, 2019):

- **Modelo em Estrela**
Neste modelo, as tabelas de dimensão relacionam-se diretamente com a tabela de factos, que se localiza centralmente e por isso assemelha-se a uma estrela. A tabela de factos encontra-se na terceira forma normal e as tabelas de dimensão possuem a informação desnormalizada.
- **Modelo em Floco de Neve**
As tabelas de dimensão relacionam-se com a tabela de factos, no entanto existem algumas que se relacionam entre si, possuindo a informação normalizada.
- **Modelo em Galáxia**
É composta por duas ou mais tabelas de factos e partilham entre si tabelas de dimensão.

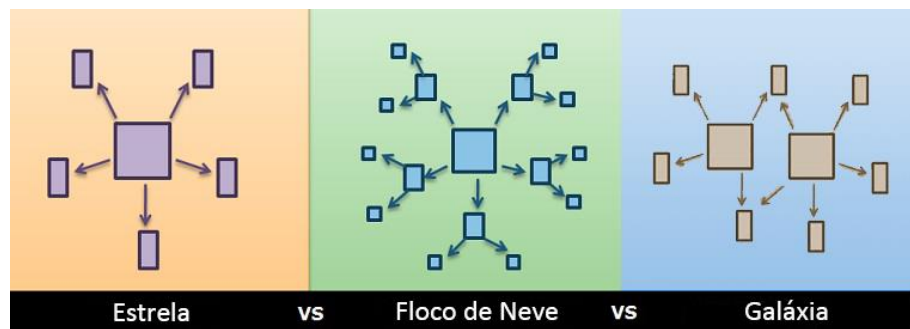


Figura 4 - Tipos de Modelos Dimensionais

Comparando os modelos, o modelo em estrela é o mais simples de utilizar. Como existem ligações diretas entre as tabelas de dimensão e de factos, os softwares conseguem aceder aos dados mais rapidamente, enquanto que nos restantes modelos são necessárias mais instruções SQL, o que torna o acesso mais lento.

Apesar disto, pelo facto de o modelo em estrela possuir a informação desnormalizada, existe espaço que está a ser desperdiçado, pois existem atributos que se repetem ao longo da mesma tabela. Mesmo assim, isto constituiu uma pequena percentagem de desperdício de memória, enquanto que os restantes modelos tornam-se mais complexos porque acrescentam tabelas ao modelo (Brito, 2015).

3.2 OLAP

O conceito OLAP refere-se a um conjunto de ferramentas que se focam em análise ad hoc de dados e que facilitam consultas de Business Intelligence. Este tipo de ferramentas ajudam na tomada de decisão, uma vez que conseguem fazer consultas aos dados dentro de um Data Warehouse complexo, rapidamente e obter a informação de forma sumariada (Araújo, 2007) .

Esta metodologia assenta numa estrutura designada de “Cubo OLAP”, que permite visualizar os dados segundo diferentes perspetivas, através da permuta de informação entre os diferentes eixos que o constituem. A Figura 5 representa o cubo multidimensional, que relaciona a informação segundo três dimensões (Região, Produto e Dia em que o produto foi adquirido).

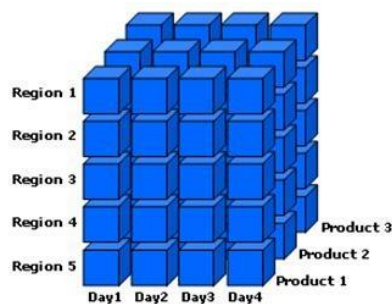


Figura 5 - Cubo OLAP (Microsoft, 2012)

Existe um conjunto de operações que podem ser aplicadas sobre o cubo para se obter a informação pretendida. Estas operações atuam como filtros e são as seguintes:

- **drill across** - Ocorre quando o utilizador realiza análises numa determinada dimensão, podendo mudar de nível (Ex: Passa de ano para mês na dimensão tempo).
- **drill down** – ocorre quando o utilizador pretende aumentar o nível de detalhe de informação. Para tal a granularidade diminui.
- **drill up** – é o contrário Drill Down, o utilizador aumenta a granularidade para diminuir o nível de detalhe da informação.
- **drill through** – o utilizador passa de uma dimensão para outra
- **slice & dice** – permite dividir a informação em partes menores, para que se consiga examinar a mesma em diferentes perspetivas e entendê-la melhor.

O armazenamento de dados pode ser efetuado de acordo com os diferentes tipos de estruturas: MOLAP, ROLAP e HOLAP (1keydata, 2019).

O primeiro, MOLAP (Multidimensional Online Analytical Processing) armazena a informação num cubo multidimensional para que os dados sejam acedidos rapidamente, e como tal pode-se executar cálculos complexos devido ao seu desempenho. O segundo, o ROLAP (Relational Online Analytical Processing) manipula diretamente os dados que se encontram armazenados numa base de dados relacional.

Por fim, o HOLAP (Hybrid Online Analytical Processing) combina o melhor do MOLAP e do ROLAP. Assim, utiliza o cubo multidimensional para obter um desempenho mais rápido e os dados relacionais para obter mais detalhe quando necessário.

3.2.1 Sistema OLAP vs. Sistema OLTP

Em Business Intelligence (BI) é comum utilizarem-se as siglas OLTP e OLAP para definir conceitos que são aplicados em contextos diferentes.

Enquanto que OLTP se refere a sistemas transacionais que processam informação que é gerada diariamente pelos sistemas de informação das organizações, os sistemas OLAP dizem respeito

a sistemas com capacidade de analisar grandes volumes de informação, para serem analisados em diferentes perspectivas (Elias, 2014).

A Tabela 9 evidencia as diferenças entre sistemas OLAP e OLTP.

Tabela 9 -Diferenças entre OLAP e OLTP (TechDifferences, 2016)

| Base de Comparação | OLTP | OLAP |
|--------------------|--|--|
| Foco | Aplica operações CRUD em bases de dados | Extraí dados para apoiar na tomada de decisão |
| Dados | Sistemas operacionais | Sistemas do tipo dimensional e/ou relacional |
| Transação | Transações curtas | Transações longas |
| Tempo | Processamento de transação menor | Processamento de transação maior |
| Consultas | Simplex | Complexas |
| Normalização | Tabelas Normalizadas | Tabelas Desnormalizadas |
| Integridade | Mantém restrições de integridade | Integridade não é afetada |
| Conteúdo dos dados | Os dados representam estado atual do negócio | Os dados apresentam várias atividades de negócio |

3.3 Ferramentas

Ferramentas de Business Intelligence permitem recolher, analisar e converter dados em *insights* valiosos para as organizações, sejam elas de pequeno, médio ou grande porte.

É através deste tipo de ferramentas que as organizações conseguem medir o seu KPI (Key Performance Indicator), que constitui um indicador-chave de desempenho relevante para determinar estratégias e obter sucesso.

A escolha do software é um fator importante, pois é através do mesmo que as empresas obtêm informações acerca do seu crescimento, visualizam as tendências e comportamentos relativos aos seus clientes. De certa forma, é através da ferramenta que se efetua a tomada de decisão.

Como as características dos softwares que existem de Business Intelligence variam, e nem todas as empresas têm as mesmas necessidades é necessário interpretar as suas características, comparando as vantagens e desvantagens que elas podem trazer.

A comparação entre ferramentas é efetuada, tendo em conta os critérios definidos no Quadrante Mágico 2018 pela Gartner (Gartner, 2018a).

Gartner é uma empresa fundada em 1970, por Gideon Gartner e que atua no ramo de prospeções sobre o mercado de IT. Encontra-se sediada no estado Kentucky, nos Estados Unidos e posiciona-se no mercado com o objetivo de apoiar empreendedores da tecnologia (OpServices, 2017).

O quadrante Mágico representa graficamente o mercado de IT num determinado período de tempo. Assim sendo, as empresas podem encontrar-se nos seguintes quadrantes:

- **Líderes** – Aqui enquadram-se as empresas que possuem tecnologias avançadas e de visão de mercado.
- **Desafiadores** – Neste quadrante encontram-se as empresas que conquistam mercado e que possuem capacidade de executar grandes projetos.
- **Visionários**- Apesar de possuírem boas capacidades de pesquisa e desenvolvimento, não têm grande poder tecnológico.
- **Concorrentes de Nicho**- Empresas que possuem um foco num determinado segmento do mercado.

3.3.1 Ferramentas de ETL

O quadrante mágico é aplicado na comparação das ferramentas ETL existentes atualmente no mercado, como se pode ver na Figura 6.



Figura 6 - Quadrante Mágico das ferramentas ETL (Gartner, 2018b)

Segundo Gartner existem alguns critérios que devem ser tidos em conta quando se faz a escolha de uma ferramenta de integração de dados. Como por exemplo (Gartner, 2018b):

- **Interoperabilidade com integração de aplicações numa só solução** – Inclui serviços como captura de dados alterados, leitura de serviços de mensagens e receber fluxos de dados (data streaming).
- **Integração de dados combinados** – Permitem a integração de dados vindos de várias fontes como aplicações, serviços e armazenamento quer na nuvem como localmente.
- **Suporte a entrega e acesso a dados a várias fontes** – Fornecer um grande leque de formas de armazenamento de dados e repositórios tais como: soluções de dados distribuídos e repositórios de gestão de dados analíticos.
- **Consistência de dados entre aplicações operacionais** – As aplicações de integração de dados devem garantir a consistência das bases de dados por todas as aplicações.

A seguir encontram-se alguns exemplos de ferramentas que podem ser utilizados para a extração, transformação e integração dos dados:

- **Talend**

O Talend é uma ferramenta de integração de dados que inclui qualidade de dados e ferramentas de gestão de API. Um dos pontos importantes na utilização do Talend é a configurabilidade e flexibilidade na criação de processos de integração de dados (Gartner, 2018b).

Oferece capacidades de qualidade de dados, assim como ferramentas de gestão de APIs que permitem melhorar a gestão dos dados e da infraestrutura da aplicação. Gartner classifica o Talend como estando no quadrante “Líderes”.

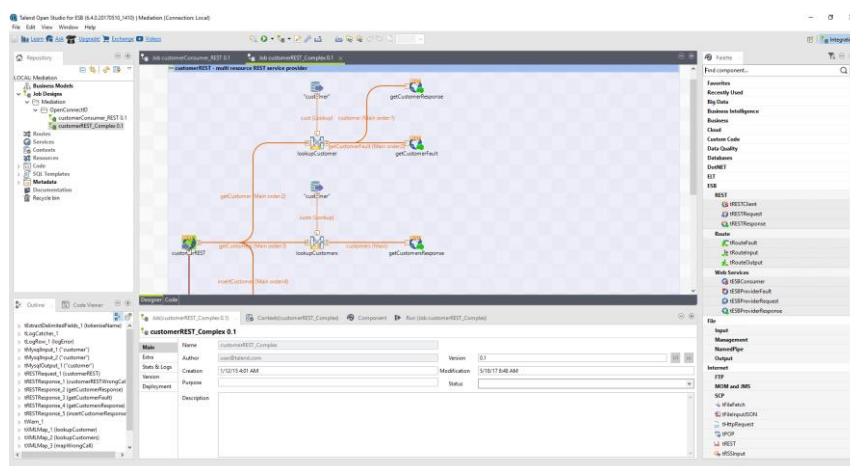


Figura 7 - Interface do Talend (Talend)

- **Microsoft SQL Server Integration Services (SSIS)**

A Microsoft oferece serviços de integração de dados através do SSIS que permite resolver grandes problemas empresariais através da cópia de dados, carregamento de dados para data

warehouses, assim como limpeza e mineração de dados. O SSIS fornece um conjunto de tarefas e transformações integradas através de ferramentas visuais (Microsoft, 2018).

Um dos pontos fortes do SSIS é a rapidez da implementação, a facilidade de uso e a capacidade de fácil integração com outras ferramentas da Microsoft como o SQL Server e os serviços do Azure na Cloud. Também inclui capacidades de combinar o uso de ETL com fluxos de trabalho corporativos e preparação dos dados (Gartner, 2018b). A Microsoft é considerada no quadrante mágico como um “Desafiador”.

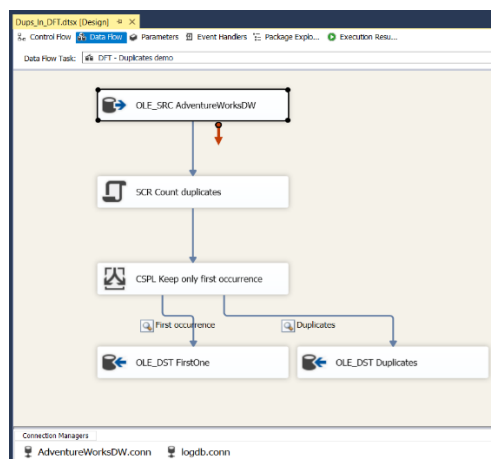


Figura 8 - Interface do SSIS (Michiels, Hans)

- **Oracle Data Integrator (ODI)**

O Oracle Data Integrator é uma plataforma de integração abrangente que permite elevados volumes de dados com grande performance. Pode ser executada como uma solução única ou integrada com outras tecnologias e aplicações da Oracle (Oracle, 2019).

A Oracle inclui ferramentas de qualidade de dados e aplicações empresariais. Também suporta SQL de Big Data, assim como streaming na nuvem e localmente (Gartner, 2018b). Gartner considera a Oracle como um “Líder” no mercado das ferramentas de ETL.

3.3.2 Ferramentas de análise de dados

Segundo Gartner, existem 15 critérios que são agrupados de acordo com cinco categorias: Infraestrutura, gestão de dados, análises e criação de conteúdo, partilha de descobertas e capacidades gerais do software.

No contexto do projeto a desenvolver, apenas foram considerados como relevantes 11 dos 15 critérios. Os critérios selecionados vão apoiar a tomada de decisão sobre ferramentas BI de análise de dados a serem utilizadas na implementação do projeto (Gartner, 2018a).

Abaixo encontram-se os critérios de seleção (Gartner, 2018a):

- **Administração, Segurança e Arquitetura de Plataformas de BI**
Possui recursos que garantem a segurança, gestão de utilizadores e utilização da plataforma. Oferecem melhor desempenho, alta disponibilidade e capacidade de recuperar dados em caso de desastre.
- **Cloud BI** – Possui capacidade de criar e gerir aplicações analíticas na nuvem e localmente.
- **Conectividade a fontes de dados**- permite que haja conexão entre os utilizadores e os dados, que se encontram em plataformas de armazenamento localmente ou na nuvem.
- **ETL e armazenamento de dados**- Ferramentas que permitem recolher, transformar e carregar dados para um ambiente independente.
- **Preparação de dados self-service** – permite ao utilizador combinar dados de diferentes fontes e criar modelos analíticos arrastando os mesmos para o painel.
- **Dashboards** – criação de painéis e exploração visual de conteúdo para se efetuarem análises à informação.
- **Descoberta de conhecimento nos dados** – permite encontrar correlações, exceções, clusters e previsões nos dados que são transmitidos visualmente.
- **Exploração Móvel** – permite visualização do conteúdo em dispositivos móveis.
- **Incorporação de Conteúdo**- contém recursos que permitem incorporar dados em aplicações através de APIs.
- **Partilha de Conteúdo** – permitem aos utilizadores publicar, partilhar e até mesmo colaborar no conteúdo analítico.
- **Facilidade de uso** – facilidade em criar conteúdo, interagir com o mesmo.

Após a aplicação dos critérios anteriormente definidos, as ferramentas são divididas em quatro quadrantes que definem a sua posição atual do mercado da análise de dados. Essa representação pode ser vista na Figura 10.



Figura 10 - Quadrante Mágico de Business Intelligence (Gartner, 2018a).

As ferramentas que foram consideradas para a análise comparativa são:

- **Sisense**

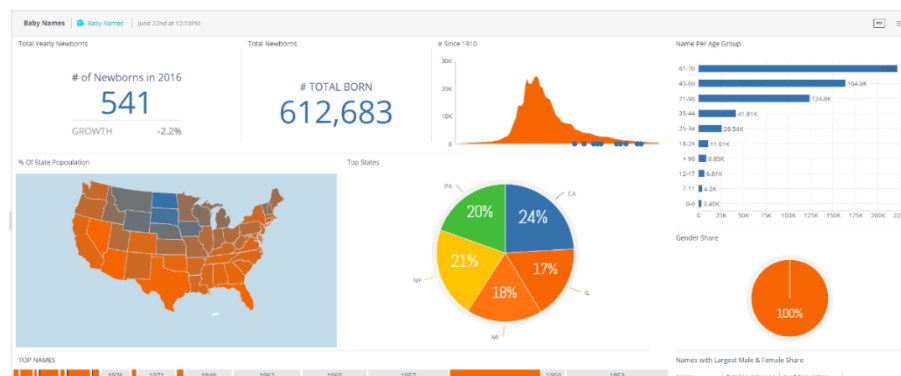


Figura 11 - Interface do Sisense (Sisense)

O Sisense é uma plataforma de Business Intelligence com um back-end que permite que utilizadores sem capacidades técnicas consigam analisar elevados conjuntos de dados de várias fontes. Também inclui um front-end para uma fácil criação de dashboards e relatórios (G2 Crowd, 2019).

Em 2017 foi adicionado melhor suporte para cloud, mobile e incorporação de conteúdo, assim como uma melhor preparação de dados “Self-service”. Foi também introduzida uma interface totalmente baseada em web e capacidades de machine learning (Gartner, 2018a).

Segundo Gartner, o Sisense fica no quadrante “Visionários”.

Qlik Sense

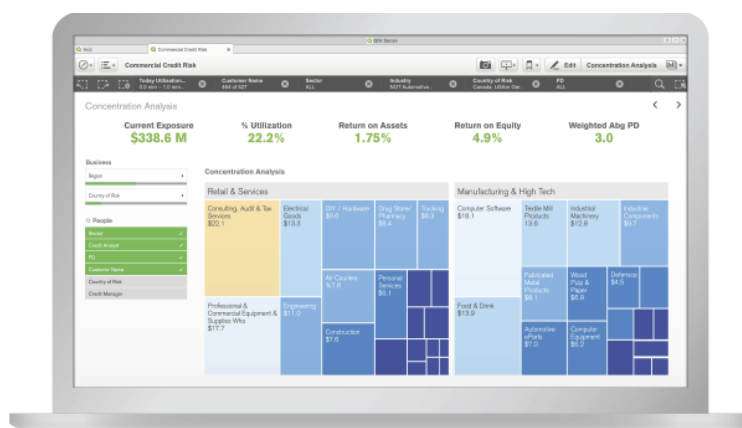


Figura 12 - Interface do Qlik Sense (Qlik)

O Qlik Sense é uma plataforma de análise de dados completa que é utilizada muitas vezes com um data mart em vez do tradicional data warehouse. Isto porque permite a receção de dados de múltiplas fontes de dados, assim como a geração de modelos complexos de dados e cálculos (Gartner, 2018a).

No entanto, a principal característica é a preparação de dados “Self-Service”, assim como as dashboards interativas. Para além disso, a Qlik tem investido em capacidades point-to-click para carregamento de dados para facilitar o desenvolvimento (Gartner, 2018a), oferece também suporte a APIs abertas e ferramentas de desenvolvimento para incorporação de dados.

O Qlik Sense também permite a criação de dashboards para todos os tipos de dispositivo (mobile e desktop), a criação de reports em vários formatos populares como Office e PDFs e incorporação com a cloud (Qlik Sense, 2019).

Segundo Gartner, o Qlik Sense situa-se no quadrante “Líderes”.

Tableau



Figura 13 - Interface do Tableau (Tableau)

O Tableau oferece um sistema analítico baseado em web que permite que qualquer utilizador mesmo sem conhecimentos técnicos consiga aceder, analisar e apresentar descobertas nos seus dados (Gartner, 2018a).

Inclui conectividade para praticamente todos os tipos de fontes de dados, uma interface baseada em “Drag-and-Drop” que permite visualizar quaisquer dados, explorar diferentes vistas e combinar várias bases de dados (Tableau, 2019a).

Permite acesso colaborativo e partilhado para todos os dispositivos através da Cloud com dashboards públicas disponíveis a partir de qualquer lugar. Também permite integração de conteúdo através de APIs e sincronização com software de negócio como Salesforce e Microsoft SharePoint (Tableau, 2019b).

Gartner posiciona o Tableau no quadrante “Líderes”.

Microsoft PowerBI

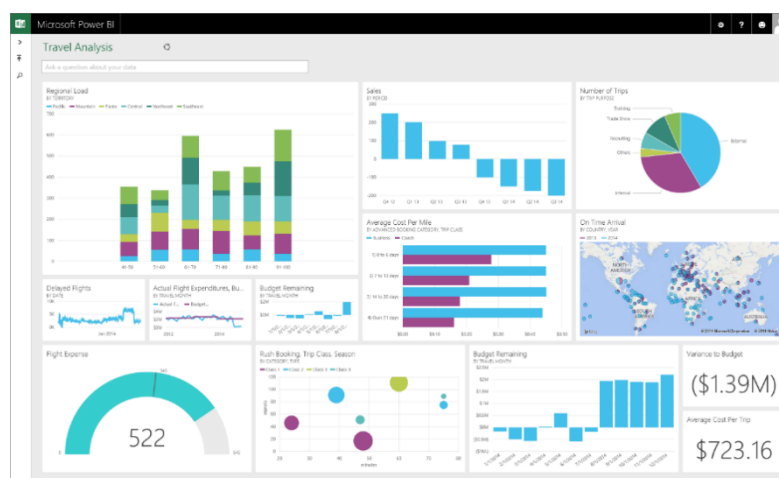


Figura 14 - Interface do PowerBI (Microsoft)

O PowerBI da Microsoft é um conjunto de ferramentas analíticas que permitem a descoberta de dados e criação de dashboards interativas. Também é possível ligação a várias fontes de dados com preparação de dados e criação de relatórios em poucos minutos, assim como criação de visualizações com ajuda de Inteligência Artificial a partir de dados já existentes (Microsoft, 2019c).

Inclui publicação e partilha de conteúdo com integração com o Office e outras ferramentas permitindo a distribuição de informações por qualquer dispositivo com poucos cliques.

Permite integração dos conteúdos do PowerBI em sites, portais e outras aplicações com vários tipos de APIs e com o kit de desenvolvimento disponibilizado.

Segundo Gartner, o Microsoft posiciona-se quadrante “Líderes”.

Looker



Figura 15 - Interface do Looker (Looker)

O Looker é uma ferramenta moderna de Business Intelligence baseada principalmente na Cloud que permite integrar, explorar e visualizar dados. Suporta vários tipos de fontes de dados e visualizações e pode ser integrada em portais e aplicações (Gartner, 2018a).

Possui uma linguagem de modelação de dados baseada em SQL chamada de LookML que os analistas de dados podem utilizar para definir métricas e dimensões de forma centralizada (Looker, 2019). Sendo a principal componente de utilização uma linguagem de programação, significa que não é a ferramenta mais fácil de utilizar e são necessários alguns conhecimentos técnicos.

Segundo Gartner, Looker posiciona-se no quadrante “Operador de Nicho”

IBM Cognos Analytics

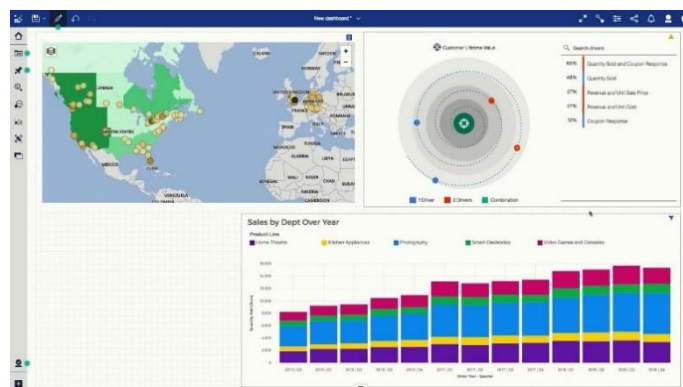


Figura 16 - Interface do IBM Cognos Analytics (IBM)

O IBM Cognos Analytics é uma ferramenta de BI que incorpora Inteligência Artificial e Machine Learning que permite descobrir informação oculta e menos óbvia utilizando detecção de padrões. Também permite a redução do tempo necessário à preparação dos dados através de automação e da inteligência integrada (IBM, 2019).

Apesar de incluir capacidades de preparação de dados “self-service” e interação visual é bastante limitado na complexidade da análise. O mesmo acontece com as capacidades móveis da ferramenta (Gartner, 2018a).

Gartner posiciona IBM no quadrante “Visionários”.

Domo

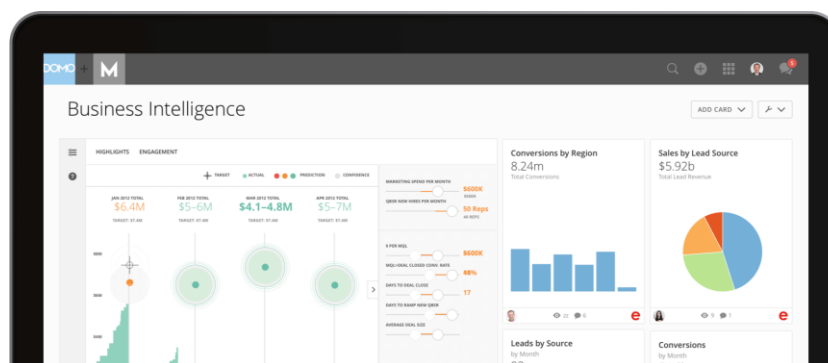


Figura 17 - Interface do Domo (Domo)

Domo é uma plataforma analítica baseada na Cloud cujo público-alvo são utilizadores do negócio que precisam de dashboards intuitivas. É utilizado geralmente em ambientes de negócio com muito pouco ou nenhum apoio de pessoal técnico de TI (Gartner, 2018a).

Devido à sua natureza totalmente baseada na Cloud permite que equipas partilhem dados em tempo real e de forma rápida e em qualquer dispositivo (Domo, 2019). Apesar de direccionada

a utilizadores não técnicos também inclui APIs para integração em outras aplicações (Domo, 2019).

Segundo Gartner, Domo posiciona-se no quadrante “Operador de Nicho”.

A Tabela 11 representa a comparação entre ferramentas de análise de dados, de acordo com os critérios de Gartner e as características que apresentam.

Tabela 11 - Comparação entre ferramentas de BI

| Ferramentas Critérios | Sisense | Qlik | Tableau | Power BI | Looker | IBM Cognos | Domo |
|--------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Administração de Plataformas de BI | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Cloud BI | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Conectividade a Fontes de Dados | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| ETL e armazenamento de dados | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | | | | |
| Preparação de dados “Self-Service” | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Dashboards | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Descoberta de conhecimento nos dados | <input checked="" type="checkbox"/> | | | <input checked="" type="checkbox"/> | | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Exploração Móvel | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Incorporação de Conteúdo | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Partilha de Conteúdo | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | | <input checked="" type="checkbox"/> |
| Facilidade de Uso | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | | | |

De acordo com a tabela verifica-se que as ferramentas de Business Intelligence estudadas apresentam características semelhantes. Existem algumas que não possuem todas as

características propostas, no entanto podem ser utilizadas pois complementam-se com ferramentas de ETL.

3.4 BI na área da Saúde

A saúde é uma das áreas em que sistemas de Business Intelligence tem vindo a ser implementados.

Comparando este sector com outros, a utilização de soluções tecnológicas nesta área é ainda recente. A introdução de soluções tecnológicas, permitiu que os sistemas tradicionais fossem agregados, para que fosse mais fácil a gestão de dados para auxiliar na tomada de decisão (Oliveira and Alves, 2012).

Integrar as diversas aplicações de saúde em apenas um único sistema, torna-se vantajoso pois consegue-se aceder a todos os dados dos utentes e evitar que erros humanos aconteçam.

Os sistemas de BI permitem cruzar dados clínicos, administrativos, obter conhecimento sobre o negócio, reduzir custos e melhorar processos (Vogt, 2013). Desta forma, muitos recursos que se encontravam em determinados setores podem ser alocados a outros, já que com a implementação de sistemas de BI os procedimentos que eram praticados sofrem alterações. O objetivo destes é automatizar tarefas, aumentar a performance das consultas aos dados dos utentes e reduzir custos (MediLab, 2017).

Especialistas em BI, defendem que existem três aspetos que podem ser conseguidos através de ferramentas de BI em organizações de saúde (SaúdeBusiness, 2016):

- **Comparação de Resultados**
Através de BI é possível saber o custo por paciente, número de consultas que cada um fez, realizar comparações de dados por região e até mesmo perceber a evolução do quadro de saúde dos utentes.
- **Apoio Diagnóstico**
Através do BI é possível detetar se determinado utente corre risco de sofrer uma determinada doença, com base na análise do seu histórico clínico e sabendo os seus hábitos. Desta forma, consegue-se refletir sobre o melhor tratamento a ser administrado.
- **Confiabilidade**
Os sistemas de BI oferecem soluções para problemas, garantindo a confiabilidade da informação, evitados que erros humanos de análise e de diagnóstico sejam cometidos. Para além disso, estes sistemas são bastante seguros, dado serem contra acessos não autorizados e a informação neles contida encontra-se de acordo com a proteção de dados, na medida em que os dados se encontram anonimizados não comprometendo a identidade do utente.

Os profissionais de saúde consideram que ferramentas de BI se tornaram grandes aliados no seu dia a dia, pois conseguem a partir das mesmas aumentar a qualidade do atendimento.

A gestão das organizações hospitalares também pode ser efetuada por processos de BI. Através de dashboards e relatórios pode-se visualizar a informação em tempo real, o que possibilita que sejam detetadas falhas atempadamente (Vogt, 2013) e evitadas crises no Sistema Nacional de Saúde e também obter informação para um determinado período de tempo.

4 Análise de valor

Este capítulo tem como objetivo analisar e utilizar técnicas para perceber se a ideia a desenvolver tem potencial para criar valor para o negócio onde se insere.

É através da análise de valor que se percebe se o caso de estudo pode ser de sucesso ou não e se justifica, muitas das vezes a existência de determinada organização.

As organizações só devem existir caso estas conseguirem criar valor. Nos casos em que o valor não exista, não existe qualquer motivo para a sua existência.

4.1 Inovar/Desenvolver Valor

As organizações devem procurar constantemente inovar os seus produtos ou até mesmo criar novos, que tragam à sociedade valor acrescentado. A inovação está geralmente associada a outros conceitos como tecnologia, criatividade e mudança. Na área de negócios, a inovação permite alterar a conduta e os resultados da organização, através da análise de fatores internos e externos (Damanpour, 2017).

O modelo de conceito de desenvolvimento apresentado por Peter Koen surge a partir do processo de inovação, essencial na inovação e criação de novos produtos e o mesmo encontra-se estruturado em três áreas:

- Fuzzy Front-End (FFE)
- Processo de desenvolvimento de novos Produtos (NPD)
- Comercialização

A FFE é a principal parte do processo de inovação e esta permite melhorar o mesmo, “aumentando valor, quantidade e probabilidade de sucesso no desenvolvimento e na comercialização de produtos” (A.Koen and M.Ajamian, 2002).

A falta de conhecimento sobre técnicas FFE, levou a que não se conseguisse comparar as mesmas entre empresas. De forma a colmatar este problema, foi desenvolvido um modelo teórico, designado de DNT cujo objetivo consiste em fornecer uma terminologia comum para FFE.

Após esta fase, procederam-se a análises aos dados de experiências efetuadas em várias organizações. Foi então que através de semelhanças e diferenças nos dados se percebeu que uma estrutura sequencial não se adequava ao modelo de desenvolvimento.

A ideia de estrutura sequencial foi abandonada e criou-se um novo conceito de desenvolvimento (NCD) que se divide em três partes:

- Motor ou olho de boi, constituído pelos elementos da organização que influenciam as atividades de forma direta ou indireta. Esta parte encontra-se presente no raio interno.
- Ainda no raio interno, encontra-se cinco elementos de atividades (Identificação de oportunidade, análise de oportunidade, geração de ideia, seleção de ideia, conceito de definição) da FFE.
- Fatores que influenciam de forma indireta o desenvolvimento de uma ideia. Estes fatores são difíceis de prever ou até mesmo controlar pela organização.

O modelo NCD, nomeadamente os cinco elementos de atividade controláveis são aplicados ao contexto do problema em questão, sendo que os resultados obtidos são apresentados nas secções seguintes.

4.1.1 Identificação da oportunidade

A oportunidade surge da necessidade de melhorar o processo de obtenção de indicadores relativamente a Dídivas de Sangue. De entre as quantidades diárias de informação recolhidas pelas diferentes entidades, muitas delas não são utilizadas para melhorar os cuidados de saúde, pelo que surgiu a oportunidade de colmatar este problema.

4.1.2 Análise da oportunidade

De forma a sustentar a oportunidade acima mencionada, é necessário efetuar a sua análise para se medir as vantagens e desvantagens inerentes à mesma.

Para efetuar o diagnóstico estratégico recorre-se à análise SWOT para identificar os fatores internos e externos que ajudam a definir os objetivos para a organização. “As letras SWOT referem-se a *Strenghts* (pontos fortes), *Weaknesses* (pontos fracos), *Opportunities* (oportunidades) e *Threats* (ameaças)” (Economias, 2017). Os pontos fortes e fracos referem-se aos fatores internos da organização, enquanto que as oportunidades e as ameaças referem-se aos fatores externos, que não são controláveis pela organização e que através da análise vão sendo detetadas. A Figura 18 refere-se à análise SWOT inerente à oportunidade identificada.



Figura 18 - Análise SWOT

4.1.3 Geração de ideia e Enriquecimento

Como forma de solucionar o problema encontrado, surge a ideia de implementar um novo sistema que utilize ferramentas de ETL e de BI para conseguir apresentar aos profissionais de saúde a informação de forma intuitiva.

4.1.4 Seleção da ideia

A ideia apresentada no ponto anterior constitui uma solução de Business Intelligence. Através de tecnologias eficazes é construído um sistema que permite obter a informação detalhada, de forma resumida e de fácil interpretação, uma vez que utiliza funcionalidades disponíveis por ferramentas, tais como tabelas, gráficos e relatórios.

4.1.5 Definição do Conceito

Através do sistema explicado anteriormente, a informação surge de forma resumida, fácil de interpretar porque através das funcionalidades disponíveis pelas ferramentas é possível obter indicadores sob a forma de dashboards e relatórios.

4.2 Análise de valor

É através da análise de valor que se consegue entender se o tema selecionado é vantajoso e traz benefícios não só para a organização, mas também para os clientes que pretendem ter acesso ao produto ou serviço.

4.2.1 Valor, valor para o cliente e valor percebido

As organizações geralmente apresentam um conjunto de valores os quais são defendidos para moldar a cultura da mesma. Apesar disto, o conceito é bastante subjetivo, porque os valores defendidos pela organização podem diferir dos valores que os clientes pretendem. De certa forma, traduz-se não só em valores monetários, como também em benefícios que os produtos e serviços comercializados oferecem e depende das necessidades que cada um tem.

Pode-se definir o conceito de valor para o cliente como “percepção do que produto ou serviço vale para um cliente versus as alternativas possíveis” (Cain, 2019). De certa forma, o cliente paga para ter em troca outros benefícios, tais como tempo, esforço, energia, qualidade. Apesar disto, o valor percebido varia entre clientes. Cada cliente avalia o custo-benefício dos produtos ou serviços e entende se o que é cobrado em troca dos benefícios é justo.

Apostar na diferenciação, faz com que os produtos/serviços se destaquem dos oferecidos pela concorrência, destacando as qualidades, oferecendo melhores funcionalidades e variedades (Marques, 2017).

No sistema desenvolvido, os benefícios superam os custos de implementação, na medida em que os profissionais de saúde possam utilizar a mesma para tomar decisões ponderadas no menor espaço de tempo.

4.3 Proposta de Valor

Segundo Alexander Osterwalder, pode-se definir o conceito de proposta de valor como “o conjunto de produtos e serviços que criam valor para um segmento específico de clientes” (Gabry, 2016).

De certa forma, a proposta de valor descreve os benefícios que são entregues aos clientes ou até mesmo aos funcionários e acionistas. Assim sendo, as organizações apresentam os motivos que as diferenciam da concorrência e a razão pela qual os clientes devem dar preferência aos seus produtos.

Para se criar valor é necessário inovar e recriar, pois as necessidades dos clientes são cada vez exigentes e mudam constantemente (Gabry, 2016).

No contexto atual, pretende-se desenvolver um sistema baseado em Business Intelligence que permita obter indicadores acerca de dívidas de sangue de uma organização. Através desta

solução, os profissionais de saúde conseguem obter a informação de que necessitam num único local, o que possibilita decisões por partes destes eficientes. Para os utentes, o sistema também se torna benéfico indiretamente, pois a decisão dos profissionais influencia os seus tratamentos e por vezes em situações de risco pode ser uma mais valia.

As tecnologias utilizadas para implementar a solução permitem que o processo garanta a qualidade da informação e a transmita ao utilizador final com resultados confiáveis.

4.4 Modelo de Negócio CANVAS

O modelo CANVAS permite estruturar facilmente o negócio e visualizar o mesmo em apenas uma única página. É a partir deste que se obtém a visão geral sobre o negócio e se entende como é criado valor para o cliente (Pereira, 2016).

Existem nove blocos em que o modelo CANVAS assenta e os quais devem ser preenchidos pela seguinte ordem:

- **Segmento de clientes** - são os clientes que estão dispostos a pagar pelo produto/serviço que se disponibiliza em troca do valor que é fornecido.
- **Proposta de valor**- “valor”/” benefício” que o produto ou serviço entrega ao cliente.
- **Canais**- Formas de comunicação e distribuição pelas quais a organização entrega valor ao cliente.
- **Relacionamento com os Clientes**- Estratégias adotadas pela organização para conseguir a retenção de clientes.
- **Receitas**- Determina as formas como o cliente pagará pelos benefícios recebidos. (Exemplos: Venda de produtos, assinatura, licença, leilões)
- **Recursos-Chave**: São os ativos fundamentais para que o negócio funcione, uma vez que permitem chegar até aos clientes (Exemplo: máquinas e instalações, patentes, recursos humanos, programadores)
- **Atividade-Chave** - São as atividades que a empresa deve fazer regularmente para conseguir entregar valor ao cliente. (Ex: Resolução de problemas, Gestão de aplicações, desenvolvimento de produtos)
- **Parceiros-chave**- São empresas/parcerias que contribuem para a concretização do negócio.
- **Estrutura de custos** - custos que derivam da operacionalização do desenvolvimento da solução.

A ideia de negócio que se pretende desenvolver foi apresentada num modelo de negócio CANVAS, que pode ser visualizado na Figura 19.

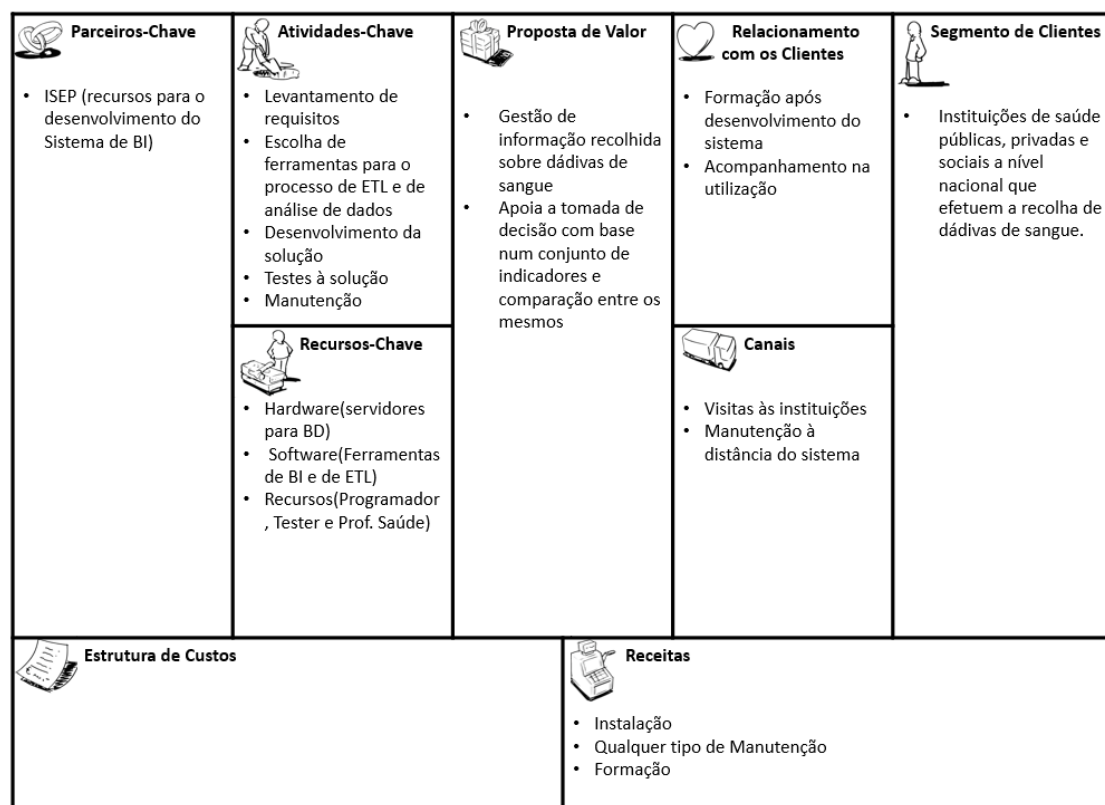


Figura 19 - Modelo de Negócio CANVAS

4.5 Cadeia de valor de Porter

Pode definir-se cadeia de valor como um conjunto de atividades que uma organização tem de realizar para criar valor para os clientes.

Porter propôs uma cadeia de valor para que as organizações examinassem as suas atividades, dividindo as mesmas em principais e de suporte.

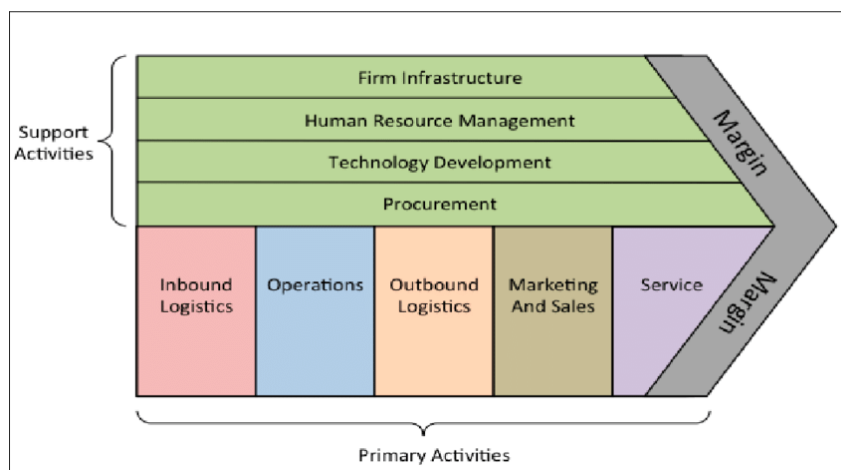


Figura 20 - Cadeia de Valor de Porter (ResearchGate.net, 2017)

As atividades principais consistem no seguinte (Mindtools, 2019):

- Logística de entrada - são atividades relacionadas com a receção, armazenamento e distribuição das entradas no produto. Relativamente à dissertação as entradas são todos registos de dados relativos a dádivas de sangue.
- Operações - estas são as atividades que transformam entradas em saídas. No contexto atual, os dados iniciais são transformados em informações úteis para o utilizador final.
- Logística de saída - essas atividades que permitem a entrega do produto ou serviço ao cliente. No caso em questão, não existe entrega de um produto físico, sendo que o que se pode considerar é apenas a manutenção do sistema.
- Marketing e vendas - são os processos que diferenciam os produtos da concorrência persuadindo os clientes a comprar os mesmos. A criação de tutoriais que mostram o modo de utilização do sistema, assim como a facilidade de utilização podem ser aplicados neste contexto.
- Serviço - são as atividades relacionadas com a manutenção do valor do produto ou serviço após a compra. Neste ponto, a existência de um serviço de helpdesk e de gestão da infraestrutura associada ao produto são essenciais.

Para além disto, as atividades de suporte das organizações encontram-se divididas em quatro áreas:

- Aquisição (compra) – recursos que uma organização necessita para operar. Neste caso, é necessário adquirir servidores para alojar bases de dados, bem como software para análise dos dados.
- Gestão de recursos humanos – É a forma como uma organização recruta e mantém os seus funcionários. Existe um conjunto de ações que incentivam os recursos a permanecerem fiéis à organização nomeadamente: promoções, bom ambiente de trabalho e estágios.
- Desenvolvimento tecnológico – atividades relacionadas com a gestão e processamento de informações. Isto é possível através de utilização de tecnologias e ferramentas recentes no mercado e que contribuem para inovar os produtos.
- Infraestrutura - sistemas que dão suporte à empresa e permitem manter as operações diárias. Para tal a organização deve definir uma estrutura organizacional composta por diferentes departamentos (Ex: Financeiro, Recursos Humanos, Administração, Desenvolvimento).

Através da análise acima referida, é evidenciado o potencial que o tema proposto tem na criação de valor e a possível adaptabilidade a uma utilização num contexto real de uma organização.

4.6 Analytic Hierarchy Process (AHP)

Thomas Saaty, em 1980, introduziu um método multicritério designado de Analytic Hierarchy Process (AHP) para ajudar a resolver problemas complexos e apoiar na tomada de decisão. Este método permite estabelecer prioridades e dividir o problema complexo em menores de forma a conseguir obter todos aspetos, sejam eles objetivos ou subjetivos.

O problema complexo é decomposto em fatores, que por sua vez podem ainda ser decompostos até ao nível mais baixo, de forma a estabelecer relações e sintetizar a informação, comparando aspetos quantitativos e qualitativos de cada alternativa (Marins, Souza and Barros, 2019). O processo AHP pode-se dividir em três etapas, as quais são enunciadas nas secções imediatamente a seguir.

4.6.1 Divisão Hierárquica

O método AHP estrutura o problema em níveis hierárquicos, de forma a seguir compreender e avaliar o mesmo. O primeiro nível da hierarquia corresponde ao problema, o segundo aos critérios e o terceiro às alternativas. Aplicado o método AHP ao contexto atual, pode-se definir a seguinte estrutura:

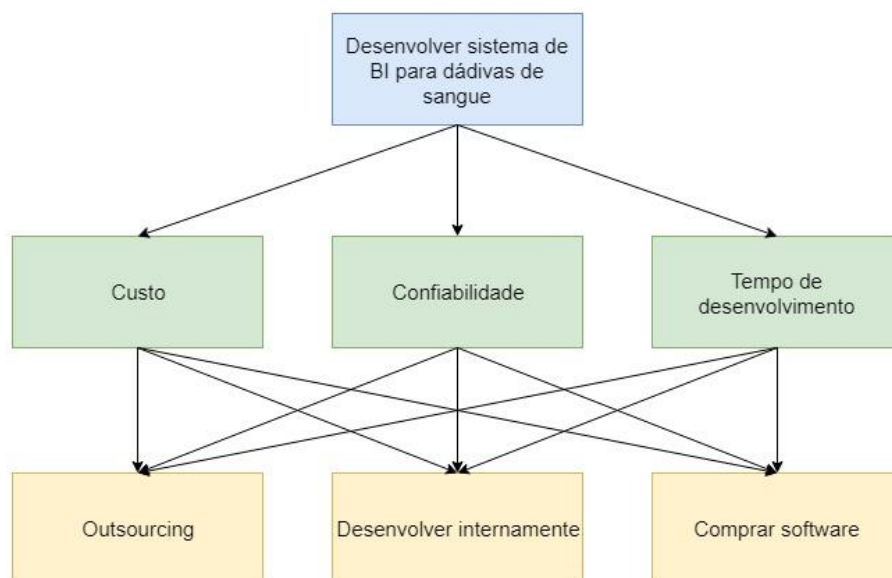


Figura 21 - Hierarquia AHP

4.6.2 Definição de Prioridades

Nesta fase são atribuídas prioridades aos critérios e alternativas apresentados na “Divisão Hierárquica”. Para tal, estabelece-se primeiro prioridades para os critérios e depois, estabelece-se prioridades para cada alternativa em relação ao critério em causa. Desta forma, consegue-se entender qual a melhor solução para determinado critério.

Para se atribuir prioridades é necessário determinar uma escala de valores que não devem exceder um total de nove fatores, para se poder comparar. Existe uma escala-padrão que foi sugerida por Saaty com o objetivo de avaliar de 1 a 9 os critérios e as alternativas no processo de decisão. Estes valores numéricos possuem uma definição qualitativa, tal como se pode visualizar na Tabela 12.

Tabela 12 - Tabela de Valores por Saaty (Santos, 2008)

| | | |
|---------|---|--|
| 1 | Igual Importância | As duas atividades contribuem igualmente para o objetivo |
| 3 | Importância pequena de uma para a outra | A experiência e o julgamento favorecem levemente uma atividade em relação à outra. |
| 5 | Importância grande ou essencial | A experiência e o julgamento favorecem fortemente uma atividade em relação à outra. |
| 7 | Importância muito grande ou demonstrada | Uma atividade é muito fortemente favorecida em relação à outra. |
| 9 | Importância absoluta | A evidência favorece uma atividade em relação à outra com o mais alto grau de certeza. |
| 2,4,6,8 | Valores intermediários | Quando se procura uma condição de compromisso entre as duas definições. |

4.6.3 Consistência Lógica

- **Prioridade relativa a cada critério**

A Tabela 13 representa as prioridades relativamente a cada critério. De acordo com a mesma é possível verificar que a confiabilidade é 8x superior ao custo, que a confiabilidade é 5x superior ao tempo de desenvolvimento e que o tempo de desenvolvimento é 3x superior ao custo.

A Tabela 14 representa os pesos finais dos critérios Custo, Confiabilidade e Tempo de desenvolvimento.

Tabela 13 - Tabela com prioridades para critérios

| Critérios | Custo | Confiabilidade | Tempo Desenvolvimento |
|------------------------------|--------------|-----------------------|------------------------------|
| Custo | 1 | 1/8 | 1/3 |
| Confiabilidade | 8 | 1 | 5 |
| Tempo Desenvolvimento | 3 | 1/5 | 1 |

Tabela 14 - Pesos dos Critérios

| Critérios | PESOS |
|------------------------------|--------------|
| Custo | 0,0768 |
| Confiabilidade | 0,7370 |
| Tempo Desenvolvimento | 0,1863 |

- **Prioridade dos critérios relativamente às alternativas**

Após definidas as prioridades dos critérios, são identificadas as prioridades de cada um dos critérios mencionados em relação às alternativas.

Tabela 15 - Critério Custo em relação às alternativas

| Custo | Outsourcing | Desenvolver internamente | Comprar Software | Pesos |
|---------------------------------|--------------------|---------------------------------|-------------------------|--------------|
| Outsourcing | 4/7 | 1/3 | 5/8 | 0,5222 |
| Desenvolver internamente | 1/7 | 0,0909 | 0,0526 | 0,0955 |
| Comprar Software | 2/7 | 1/2 | 1/3 | 0,3823 |

Tabela 16 - Critério Confiabilidade em relação às alternativas

| Confiabilidade | Outsourcing | Desenvolver internamente | Comprar Software | Pesos |
|---------------------------------|--------------------|---------------------------------|-------------------------|--------------|
| Outsourcing | 1/5 | 1/3 | 1/5 | 0,2364 |
| Desenvolver internamente | 0,0385 | 0,0667 | 0,0816 | 0,0623 |
| Comprar Software | 3/4 | 3/5 | 3/4 | 0,7013 |

Tabela 17 - Critério Tempo Desenvolvimento em relação às alternativas

| Tempo Desenvolvimento | Outsourcing | Desenvolver internamente | Comprar Software | Pesos |
|---------------------------------|--------------------|---------------------------------|-------------------------|--------------|
| Outsourcing | 1/6 | 1/6 | 1/9 | 0,1524 |
| Desenvolver internamente | 2/3 | 5/7 | 7/9 | 0,7208 |
| Comprar Software | 1/6 | 1/9 | 1/9 | 0,1268 |

- **Avaliar a consistência das prioridades**

Após serem atribuídas todas as prioridades são realizados cálculos para entender qual a melhor decisão de acordo com os dados definidos e o ranking das alternativas.

De acordo com a Tabela 18, a melhor alternativa relativamente ao problema identificado inicialmente é “Comprar Software”.

Tabela 18 - Pesos Finais das alternativas

| | Outsourcing | Desenvolver internamente | Comprar Software | PESO |
|---------------------------------|--------------------|---------------------------------|-------------------------|-------------|
| Outsourcing | 0,5222 | 0,2364 | 0,1524 | 0,24062326 |
| Desenvolver internamente | 0,0955 | 0,0623 | 0,7208 | 0,173094597 |
| Comprar Software | 0,3823 | 0,7013 | 0,1268 | 0,524546037 |

5 Design da Solução Proposta

Neste capítulo é apresentada de forma breve as regras de negócio acerca das Dádivas de Sangue, bem como toda a abordagem técnica que permite alcançar os objetivos propostos nesta dissertação. De certa forma, são fornecidas explicações acerca do domínio das dádivas de sangue, são apresentados os atores que vão interagir com o sistema, são nomeados os requisitos não funcionais e funcionais do mesmo e ainda são apresentadas propostas de arquitetura para a solução.

Para além disto, são realizadas comparações entre as arquiteturas de forma a eleger a que mais se adequa ao sistema a desenvolver e ainda é apresentada a fonte de dados utilizada e as estruturas que foram necessárias para a elaboração da arquitetura.

5.1 Domínio das Dádivas de Sangue

Antes de se idealizar a arquitetura para o sistema, é essencial perceber todo o processo acerca das Dádivas de sangue, uma vez que se trata de um processo em que a validação das mesmas é rigoroso, dado ser essencial despistar fatores de risco. Quando o dador pretende realizar a dádiva é lhe efetuado um registo. Este funciona como triagem, em que o profissional de saúde realiza ao possível dador algumas medições e questões, de forma a despistar de imediato causas que impeçam de avançar para a colheita de sangue.

É de realçar que estas entrevistas são confidenciais e nos registos o dador é identificado por um número que é único. Para além de dados pessoais (Ex: Sexo, data de nascimento, localidade postal), estes registos possuem informações sobre a instituição a qual o profissional de saúde pertence, alguns indicadores de saúde (Ex: peso, altura, Tensão Máxima, Tensão Mínima, Hemoglobina, Tipo de sangue) e as conclusões apuradas previamente junto do dador e com base nos testes efetuados. Estas conclusões permitem validar se o indivíduo se encontra apto ou não a ser um possível dador de sangue.

Posto isto, as colheitas podem ser realizadas por dadores internos (indivíduos que se registaram) nos casos em que a conclusão obtida no registo seja positiva ou por dadores externos (indivíduos que não possuem número de dador e que efetuam a dádiva quando acompanhados por declaração médica que justifique que se encontram aptos). A colheita a efetuar aos dadores não tem sempre o mesmo fim, pelo que existem diferentes tipos de dádivas. Esta pode ser total, como pode ser apenas para retirar determinados componentes sanguíneos.

Caso a conclusão obtida nas colheitas seja positiva, sucede-se a análise de sangue. Nesta última etapa, são efetuados um conjunto de testes ao sangue obtido na colheita, para verificar se os valores para os componentes sanguíneos, índices presentes no hemograma e os marcadores

serológicos para deteção de Hepatites Virais se encontram de acordo com os valores de referência, e assim, conseguir-se retirar conclusões acerca da Dádiva.

5.2 Identificação dos Atores

Profissionais de saúde - Consideram-se Enfermeiros e Médicos, os profissionais de saúde que vão interagir com o sistema. Estes contactam diretamente com o utente e necessitam de consultar o sistema para efetuarem a tomada de decisão com base nos dados apresentados.

5.3 Requisitos

A análise de requisitos permite identificar as necessidades dos diferentes atores que vão interagir com a aplicação. De certa forma, ajudam na identificação de oportunidades inerentes ao negócio e simplificam o design do sistema (ReQtest, 2018).

5.3.1 Requisitos Não Funcionais

Os requisitos não funcionais especificam como o sistema executa determinada função. No contexto atual, podem considerar-se os seguintes requisitos não funcionais:

- Desempenho
- Usabilidade
- Integridade dos dados
- Confiabilidade
- Histórico

O sistema a desenvolver deve apresentar um bom desempenho, pois pretende-se que os profissionais de saúde consigam efetuar consultas à informação dos utentes a qualquer momento.

Para além disso, é importante que o sistema seja de fácil utilização. A usabilidade é o requisito que se relaciona com o desempenho do sistema e a satisfação do cliente relativamente ao mesmo. É através deste requisito que se avalia a facilidade de aprendizagem e de utilização do sistema. O esforço exigido para aprender a utilizar a aplicação, a velocidade de execução das tarefas e a quantidade de erros no sistema, influenciam este requisito.

A integridade dos dados relaciona-se com a confiabilidade, pois como os dados em causa são informações clínicas sobre utentes, os dados devem encontrar-se anonimizados e as alterações que são efetuadas nas fontes de dados devem refletir-se na solução final.

Por fim, o sistema deve manter o histórico da informação, pois dados obtidos à posteriori poderão em determinadas situações apoiar a decisão.

5.3.2 Requisitos Funcionais

Os Requisitos funcionais permitem especificar o que o sistema deve fazer, ou seja, deve descrever o comportamento para quando determinadas condições são consideradas no sistema (Eriksson, 2015).

Podem considerar-se os seguintes requisitos funcionais:

- Filtrar informação
- Obter indicadores
- Visualizar informação através de dashboards
- Obter relatórios

Os profissionais de saúde devem conseguir obter a informação, de acordo com os filtros aplicados. Desta forma, conseguem obter um conjunto de indicadores que influenciam a tomada de decisão.

A informação obtida pode ser visualizada através de dashboards e relatórios, disponibilizadas através de ferramentas de BI, caso se pretenda arquivar a informação sobre os utentes ou até mesmo partilhar entre profissionais de saúde.

A Figura 22 representa o diagrama de casos de uso relativamente às interações dos atores com o sistema.

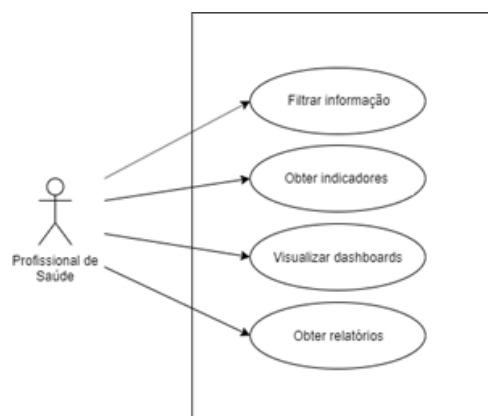


Figura 22 - Diagrama de casos de uso

Filtrar Informação

São os profissionais de saúde que se encontram envolvidos neste caso de uso. Estes, conseguem filtrar a informação que pretendem visualizar através de diferentes filtros (Ex: Ano, Mês, Dia de Semana, Sexo, Faixa Etária, Estado Civil, etc.). Desta forma, na interface de visualização a informação aumenta ou diminui de acordo com o filtro aplicado.

Obter Indicadores

Os indicadores podem ser obtidos, mesmo que não se selecione nenhum filtro. Quando tal situação acontece, na interface de visualização aparecem todos os resultados para cada um dos indicadores. De acordo com o contexto atual, existe um conjunto de indicadores que se conseguem obter ao nível da informação dos Dadores, Colheitas e Análises de Sangue.

- **Dadores**
 - Número de registos efetuados junto das Brigadas
 - Número de dadores registados
 - Informações sobre os dadores (valor mínimo, médio e máximo para peso, altura, hemoglobina, tensão máxima, tensão mínima)
 - Número de dadores por sexo
 - Número de dadores por faixa etária e sexo
 - Número de dadores por estado civil
 - Número de dadores por profissão
 - Número de dadores por localização
 - Número de dadores por grupo sanguíneo
 - Número de brigadas que efetuaram os registos
 - Número de registos por Dia de Semana (fim de semana vs. semana)
 - Número de dadores por conclusão retirada na triagem
 - Número de dadores por conclusão retirada nas colheitas
 - Número de dadores por conclusão retirada nas análises de sangue

- **Colheitas**
 - Número Total de colheitas
 - Número de colheitas que se realizaram por dadores registados
 - Número de colheitas que se realizaram por dadores não registados
 - Número de dadores registados que participaram nas colheitas
 - Número de colheitas por Hospitais, evidenciando também o número de colheitas que neles foram efetuadas por dadores registados e não registados
 - Número de colheitas por Componentes Sanguíneos retirados durante a extração das mesmas
 - Número de colheitas por grupo Sanguíneo
 - Número de colheitas por Tipo de colheitas
 - Número de colheitas segundo os motivos de inutilização

- **Análises de Sangue**
 - Número de doadores cujas análises de sangue foram analisadas
 - Número de colheitas analisadas
 - Valores mínimo, médio e máximo para cada um dos testes realizados às análises de sangue (Anti_HIV, Anti_HTLV, Anti_HCV, Anti_HBc, ALT, VDRL, WBC, Ag_HV, MCV, MCH, MCHC, HGB, NEU%, HCT, PLT, RBC)

Visualizar dashboards

Quando os profissionais de saúde acedem à aplicação, conseguem visualizar a informação através de dashboards.

Obter Relatórios

Os profissionais de saúde quando acedem à aplicação, conseguem executar relatórios com a informação que pretendem analisar.

5.4 Arquiteturas Possíveis

Para o sistema desenvolvido foi necessário idealizar possíveis arquiteturas, mesmo sem ter um conhecimento prévio da estrutura do sistema. A implementação da arquitetura varia consoantes os dados de origem, o tratamento que é necessário aplicar sobre os mesmos ou até mesmo o momento em que se pretende visualizar os resultados.

Em BI existe um conjunto de tecnologias que são utilizadas na conceção da arquitetura, de acordo com a metodologia que se pretende aplicar, nomeadamente podem ser utilizados componentes como Data Warehouse (DW), Data Mart, sistemas operacionais (ODS), Data Mining, ETL, entre outros (Elias, 2018).

No caso de os dados de origem não necessitarem de grandes limpezas e transformações, podem ser acedidos diretamente a partir dos ODS por ferramentas de BI que analisam os dados e transmitem-nos ao utilizador final. Por outro lado, caso os dados de origem necessitem de grandes transformações, podem ser efetuadas as devidas transformações antes de os dados serem apresentados para o utilizador. Desta forma, após os dados passarem pelo processo de limpeza e transformação são carregados em um data warehouse, que por sua vez possui um Data Mart com a informação das Dídivas de Sangue. As ferramentas de BI acedem diretamente ao Data Mart e só depois é que a informação é transmitida ao utilizador.

A Figura 23 e a Figura 24 constituem dois diagramas que idealizam as arquiteturas propostas para o sistema que foi desenvolvido.

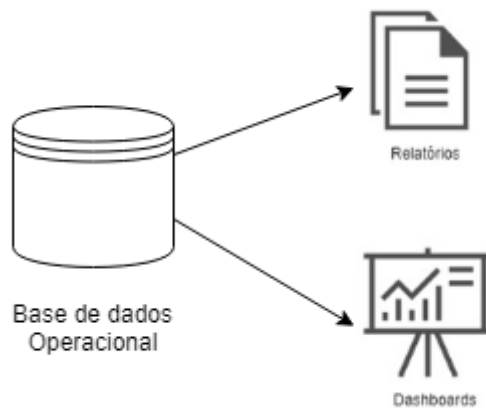


Figura 23 - Arquitetura 1

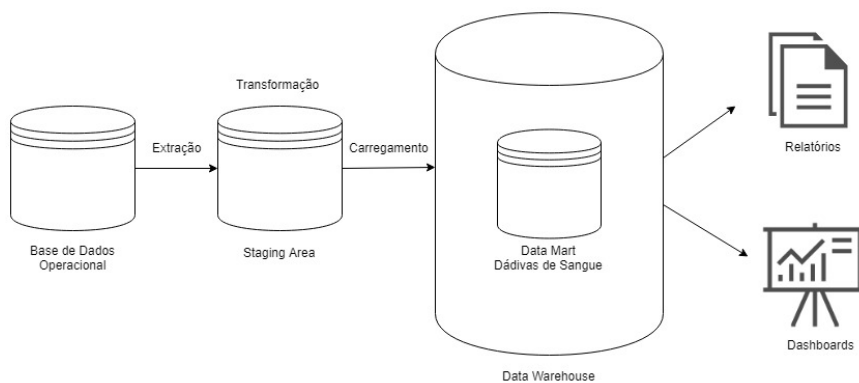


Figura 24 - Arquitetura 2

Comparando as duas arquiteturas, verifica-se que a diferença reside no repositório onde as ferramentas de BI vão aceder à informação a ser analisada. Enquanto que na Figura 23 são efetuadas queries diretamente sobre a base de dados operacional (ODS), na Figura 24 os dados só são acedidos quando se encontram no Data Mart que retrata a área de determinado negócio que se pretende analisar. A escolha das tecnologias a aplicar sobre o sistema deve ter em consideração as características de cada arquitetura.

5.5 Arquitetura Selecionada

Para a solução foi aplicada a abordagem defendida por Ralph Kimball, em que a informação relativa a Dávivas de Sangue é armazenada num Data Mart.

Seguindo esta abordagem, torna-se vantajoso a sua implementação pois caso seja necessário adicionar novas áreas à solução no futuro, a criação de novos Data Marts não comprometem o

que já existe. Para além disso, mesmo que se pretenda alterar a estrutura do sistema, quer para a adição/ remoção de funcionalidades, quer para a sua manutenção, consegue-se fazê-lo rapidamente e sem grandes problemas. O sistema a devolver deve possuir mecanismos que consigam guardar o histórico de informação, para que a decisão seja tomada com base em resultados obtidos no passado. Desta forma, foi adotada a 2ª arquitetura, apresentada na Figura 24, por esta se adequar ao conjunto de características pretendidas.

5.5.1 Fontes de dados

Os dados sobre as dádivas de sangue foram disponibilizados numa base de dados operacional Access que contém informações acerca de dadores, das colheitas realizadas junto destes, das instituições de recolha da Dádiva, dos componentes sanguíneos, do resultado de análises de sangue, conclusões obtidas e motivos de inutilização das Dádivas. Para se compreender as características desta base de dados é apresentado o respetivo modelo relacional na Figura 25.

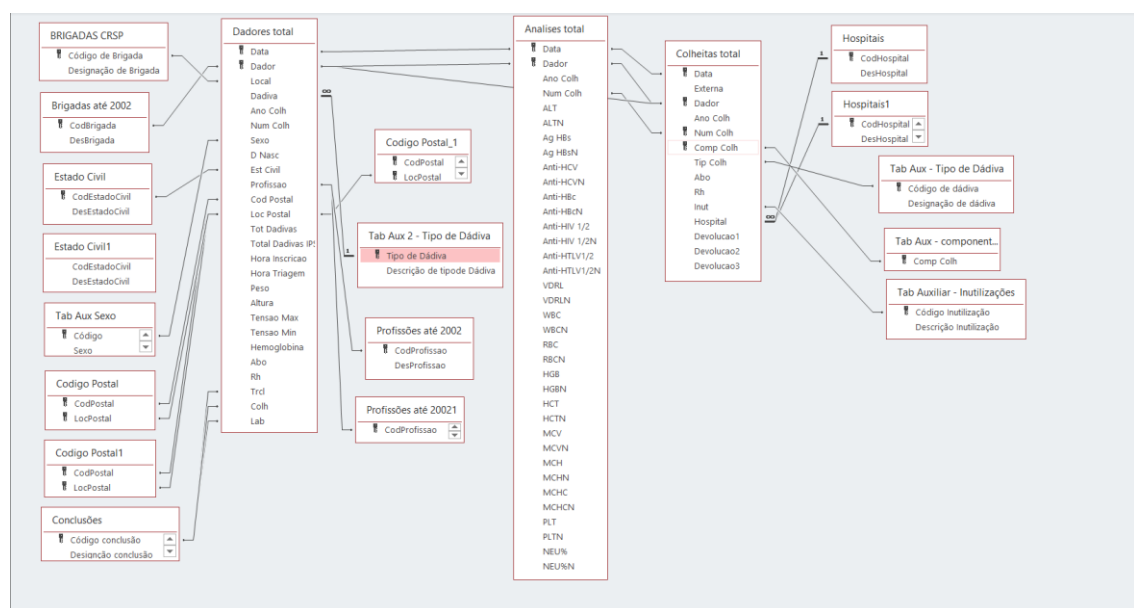


Figura 25 - Modelo de base de dados operacional

A base de dados fornecida continha uma amostra de dados que abrange informação durante um período de tempo, nomeadamente os anos de 2000, 2001 e 2002.

Após análise à mesma, verificou-se que esta possuía alguns problemas, nomeadamente tabelas com informação repetida (Ex: “Hospitais” e “Hospitais1”, “Brigadas CRSP” e “Brigadas até 2002”, “Estado Civil” e “Estado Civil1”, “Código Postal” e “Código Postal1”, “Profissões até 2002” e “Profissões até 2021”). Para além disso, verificou-se que existia a tabela “Tab Aux-Componentes” incompleta, uma vez que só continha os códigos para os componentes, não apresentando a respetiva designação.

Em relação às tabelas “Tipo Aux 2- Tipo de Dádiva” e “Tab Aux- Tipo de Dádivas”, quando comparadas, existia códigos diferentes para a mesma descrição.

Por fim, as tabelas de “Dadores total”, “Colheitas total” e “Análises total” possuem chaves compostas. A tabela “Dador” possui as chaves “Dador” e “Data”, uma vez que se verificou que para o mesmo ano existia mais do que um registo para o mesmo dador, mas a data de registo era diferente.

A tabela “Colheitas total” possui as chaves “Data”, “Dador”, “Num Colh” e “Comp Colh”, uma vez que o mesmo dador, na mesma data possuía o mesmo número de colheita, mas para componentes diferentes.

A tabela “Análises total” possui as chaves “Dador” e “Data”, uma vez que para o mesmo dador não existe mais do que uma análise na mesma data.

5.5.2 Staging Area

A Staging Area é um repositório intermédio e temporário para o qual os dados de origem são extraídos. Esta permite agilizar o processo de consolidação da informação, já que muitas vezes os dados originais são obtidos a partir de diferentes fontes, sem comprometer a integridade dos mesmos. Os dados que são transferidos para a Staging Area, sofrem um conjunto de alterações, com o objetivo de eliminar informação que se considere desnecessária e uniformizar os formatos das diversas fontes, para mais tarde poderem ser direcionados para o repositório de destino.

De certa forma, a Staging Area possui as seguintes funcionalidades (Brito, 2014):

- **Extração:** A informação é extraída das fontes de dados para a Staging Area. Esta etapa serve para diminuir os acessos ao sistema operacional, de forma a garantir que recursos desnecessários sejam consumidos.
- **Transformação:** Os dados já na Staging Area sofrem todas as transformações necessárias, incluindo a limpeza dos dados.
- **Carregamento:** Após os dados serem tratados são carregados no Data Mart.

A Figura 26 evidencia a Staging Area idealizada para o sistema desenvolvido, baseado numa base de dados Access disponibilizada pela instituição de saúde que se responsabiliza pela gestão de dádivas de sangue.

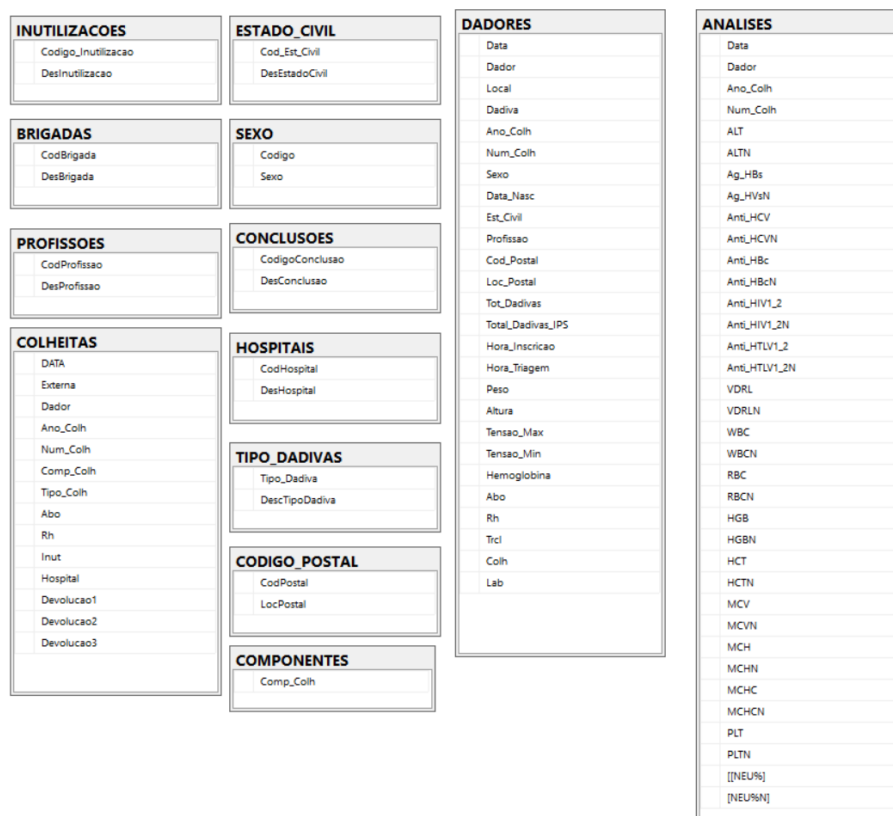


Figura 26 - Modelo da Staging Area

A Staging Area foi estruturada de modo a aproximar-se o mais possível da solução final para evitar problemas futuros de carregamento de dados para o Data Mart. Desta forma, houve necessidade de uniformizar a informação que se encontrava na base de dados Access, uma vez que nesta última se verificava que a mesma informação se encontrava repetida em várias tabelas. De acordo com a Figura 26 que retrata a Staging Area, nesta foi reduzido o número de tabelas e uniformizou-se a informação. Nas tabelas “BRIGADAS”, “CODIGO_POSTAL”, “CONCLUSOES”, “ESTADO_CIVIL”, “HOSPITAIS”, “PROFISSOES”, “INUTILIZACOES”, “SEXO”, manteve-se no seu conteúdo o código e a respetiva descrição. Na tabela “COMPONENTES”, uma vez que esta já se encontrava incompleta é caracterizada apenas pelo código, pelo que os componentes sanguíneos são identificados unicamente por estes.

As tabelas “DADORES”, “COLHEITAS” e “ANALISES” possuem no seu conteúdo códigos das tabelas com as quais se ligam. Optou-se por obter a descrição para esses códigos do lado destas, para que a informação ficasse o mais padronizada possível, evitando assim, redundância nos dados.

A tabela de “DADORES” contém a data de registo, o número do dador, a descrição da brigada, o tipo de Dádiva, o sexo, o Estado Civil, a profissão e as conclusões obtidas em cada uma das fases da dádiva (Registo, Colheita, Analise) para cada registo dos dadores. Para além disso, ainda possui o ano de registo, a data de nascimento, a hora de inscrição e de triagem, o número

total de d  divas realizadas, o peso, a altura, o grupo sangu  neo, os valores para a tens  o m  xima e m  nima e hemoglobina.

A tabela “COLHEITAS” cont  m a data da colheita, o n  mero de dador, o n  mero da colheita, o tipo de colheita, os componentes sangu  neos retirados na colheita, as inutiliza  es e o Hospital onde se efetuou a mesma.

Por fim, a tabela “ANALISES” cont  m o n  mero de dador, o n  mero da colheita, a data da respetiva an  lise e os valores para cada um dos testes efetuados   s an  lises de sangue.

5.5.3 Modela  o do DW

Nesta disserta  o    seguida a metodologia de Kimball. Um Data Warehouse    composto por um ou mais Data Marts, no entanto, no presente contexto, apenas existe um Data Mart referente ao resultado de An  lises de Sangue.

Quando se produz a modela  o de um Data Mart    necess  rio definir um conjunto de etapas que s  o essenciais na sua arquitetura, e que a seguir s  o mencionadas:

- Identificar a   rea de neg  cio
- Identificar o n  vel de granularidade/detalhe
- Identificar as dimens  es
- Identificar os factos

   a **  rea de neg  cio** que permite identificar o conte  do das informa  es que comp  em um determinado sistema numa organiza  o. No presente contexto, cada linha da tabela de Factos no Data Mart representa o resultado de An  lises Sangu  neas relativas a dadores de sangue.

O **N  vel de Granularidade** define o n  vel mais baixo que se pode obter as m  tricas na tabela de Factos, permitindo assim ao utilizador uma maior capacidade anal  tica, uma vez que a informa  o surge ao mesmo de forma detalhada. Cada linha da tabela de Factos reflete o resultado de an  lises sangu  neas efetuada a um determinado dador, numa determinada data, no   mbito de uma dada colheita.

Ap  s an  lise do sistema operacional, foram identificadas as dimens  es que d  o resposta   s necessidades dos utilizadores e ainda foram analisadas as t  cnicas de Slowly Changing Dimension que devem ser usadas para cada um dos atributos que constituem as dimens  es.

As dimens  es que se consideram necess  rias para o desenvolvimento da aplica  o s  o:

- DimDate
- DimDador
- DimColheitas
- DimJunk

DimDate

A Figura 27 representa a tabela que reflete a dimensão data. No Data Warehouse torna-se importante rastrear a informação armazenada ao longo do tempo, quer para verificação de histórico, quer para efetuar pesquisa de dados do serviço. A informação pode ser obtida a vários níveis, no entanto o valor mais baixo a considerar nesta dimensão é o dia.

| DimDate | |
|------------------|--|
| datekey | |
| fulldate | |
| year | |
| semester | |
| month | |
| monthName | |
| week | |
| DayNumberOfYear | |
| DayNumberOfMonth | |
| DayNumberOfWeek | |
| DayOfWeek | |
| Weekend | |
| trimester | |
| Quarter | |

Figura 27 - Tabela de dimensão Data (DimDate)

Na tabela “DimDate” não foi aplicada a técnica de Slowly Changing Dimension, uma vez que os atributos que a constituem referem-se a datas que não há interesse em modificar.

DimDador

Esta dimensão representa toda a informação relativa ao registo do Dador, podendo a mesmo ser visualizada na Figura 28. Nesta tabela encontram-se todos os registos dos dadores, podendo um dador ter vários registos, caso pretenda ao longo do mesmo ano efetuar a dádiva de sangue mais do que uma vez ou até em anos diferentes.

Em cada registo, encontram-se as informações apuradas pelo profissional de saúde junto dos dadores, inclusive as conclusões obtidas em cada fase da dádiva (Registo, colheita e análise de sangue), pois quando um dador vai realizar a dádiva é muitas vezes importante saber o histórico de todas as conclusões obtidas em processos anteriores. Apesar de o dador se registar, pode em qualquer altura do processo desistir e, portanto, decidir não dar sangue, não prosseguindo para a fase de Colheita.

| DimDadores |
|-------------------|
| dadorkey |
| Data |
| Dador |
| Local |
| Dadiva |
| Ano_Colh |
| Num_Colh |
| Sexo |
| Data_Nasc |
| Est_Civil |
| Profissao |
| Cod_Postal |
| Loc_Postal |
| Tot_Dadivas |
| Total_Dadivas_IPS |
| Hora_Inscricao |
| Hora_Triagem |
| Peso |
| Altura |
| Tensao_Max |
| Tensao_Min |
| Hemoglobina |
| Abo |
| Rh |
| Trcl |
| Colh |
| Lab |
| actual |
| DataEfetivo |
| DataExpiracao |

Figura 28 - Tabela de dimensão Dadores

Na tabela “DimDadores” aplicou-se o SCD do tipo 1 aos atributos “Abo”, “Altura”, “Data Nasc”, “Rh”. Estes atributos, nomeadamente o “Abo”, “Rh” e “Data Nasc” não mudam, pelo que os mesmos podem ser substituídos pelo valor mais atual. Em relação à altura, apesar de poder mudar (considerando que a fase de crescimento é até aos 20 anos) não se considera que influencie a Dádiva, pelo que interessa saber sempre o valor mais atual.

Em relação aos atributos “Ano_Colh”, “Cod_Postal”, “Colh”, “Dadiva”, “Data”, “Est_Civil”, “Hemoglobina”, “Lab”, “Loc_Postal”, “Local”, “Num_Colh”, “Peso”, “Profissao”, “Sexo”, “Tensao_Max”, “Tensao_Min”, “Tot_Dadivas”, “Total_Dadivas_IPS”, “Trcl” considerou-se que era importante manter o histórico para retirar conclusões, pelo que foi aplicado o SCD do tipo 2 a estes.

DimColheitas

Esta dimensão contém informações sobre as colheitas de sangue que são efetuadas aos dadores, sendo apresentada graficamente na Figura 29. Esta tabela possui informação acerca do dador que realizou a colheita, os componentes sanguíneos retirados na mesma, tipos de sangue, tipo de dádiva, instituição onde se efetuou a colheita e os motivos de inutilização da colheita, caso existam.

| DimColheitas | |
|---------------|--|
| colheitakey | |
| DATA | |
| Externa | |
| Dador | |
| Ano_Colh | |
| Num_Colh | |
| Comp_Colh | |
| Tipo_Colh | |
| Abo | |
| Rh | |
| Inut | |
| Hospital | |
| Devolucao1 | |
| Devolucao2 | |
| Devolucao3 | |
| actual | |
| DataEfetivo | |
| DataExpiracao | |

Figura 29 - Tabela de dimensão Colheitas

Na tabela “DimColheitas” aplicou-se o SCD do tipo 1 apenas aos atributos “Abo” e “Rh”, dado estes possuírem sempre a mesma informação. Os atributos “Ano_Colh”, “Comp_Colh”, “Dador”, “DATA”, “Devolucao1”, “Devolucao2”, “Devolucao3”, “Externa”, “Hospital”, “Inut”, “Tipo_Colh” devem ter histórico, pelo que foi aplicado a estes o SCD do tipo 2.

Após agrupar todos os atributos nas devidas dimensões, verificou-se que existiam ainda alguns que não deviam estar presentes nas dimensões retratadas anteriormente, no entanto também não se pretendia descartar essa informação, uma vez que pode ajudar a retirar conclusões. Optou-se por criar uma dimensão denominada **DimJunk** para armazenar informação não numérica relativa às análises de sangue. Esta dimensão, encontra-se representada na Figura 30.

| DimJunk | |
|---------------|--|
| junkkey | |
| Data | |
| Dador | |
| Ano_Colh | |
| Num_Colh | |
| ALT | |
| Ag_HBs | |
| Anti_HCV | |
| Anti_HBc | |
| Anti_HIV1_2 | |
| Anti_HTLV1_2 | |
| VDRL | |
| WBC | |
| RBC | |
| HGB | |
| HCT | |
| MCV | |
| MCH | |
| MCHC | |
| PLT | |
| [[NEU%] | |
| actual | |
| DataEfetivo | |
| DataExpiracao | |

Figura 30 - Tabela de dimensão Junk

Na tabela “DimJunk” foi aplicado apenas o SCD do tipo 2 a todos os atributos que representam informação não numérica relativa aos testes realizados nas análises de sangue. Esses atributos são: “NEU%”, “Ag_HBs”, “ALT”, “Ano_Colh”, “Anti_HBc”, “Anti_HCV”, “Anti_HIV1_2”, “Anti_HTLV1_2”, “Data”, “HCT”, “HGB”, “MCH”, “MCHC”, “MCV”, “PLT”, “RBC”, “VDRL”, “WBC”. Esta informação foi útil, por vezes, quando analisada em conjunto com os resultados numéricos dos testes das análises de sangue, no entanto esta informação não se refletiu diretamente nos indicadores apresentados mais tarde ao utilizador.

Após identificar as dimensões, o último passo na elaboração da arquitetura de Data Mart, segundo Kimball, é nomear os factos. Estes encontram-se na tabela de Factos, sendo representados por variáveis numéricas e referem-se aos resultados obtidos nas análises de sangue. Na construção do Data Mart foi utilizado o modelo em estrela, em que todas as dimensões se relacionam diretamente com a tabela de Factos. Este tipo de modelo para além de ser considerado o mais adequado ao utilizador torna-se também vantajoso na medida em que oferece um melhor desempenho nas consultas.

No contexto atual, a tabela de Factos denomina-se de **FactResultadosAnalises** e nesta constam todas as chaves das dimensões, bem como valores numéricos relativos a componentes sanguíneos, índices presentes no Hemograma e marcadores serológicos para se poder retirar conclusões acerca da presença do vírus da Hepatite B, C, HIV, HTLV, Sífilis e outras doenças nos dadores. A Figura 31 representa graficamente a tabela de Factos.

| FactResultadosAnalises | |
|------------------------|---------------|
| 🔑 | dateKey |
| 🔑 | colheitaKey |
| 🔑 | junkKey |
| 🔑 | dadorkey |
| | ALTN |
| | Ag_HVsN |
| | Anti_HCVN |
| | Anti_HBcN |
| | Anti_HIV1_2N |
| | Anti_HTLV1_2N |
| | VDRLN |
| | WBCN |
| | RBCN |
| | HGBN |
| | HCTN |
| | MCVN |
| | MCHN |
| | MCHCN |
| | PLTN |
| | [NEU%N] |

Figura 31 - Tabela de Factos Resultados Análises

Por fim, pode-se visualizar na Figura 32 o modelo de Data Mart implementado, de forma simplificada.

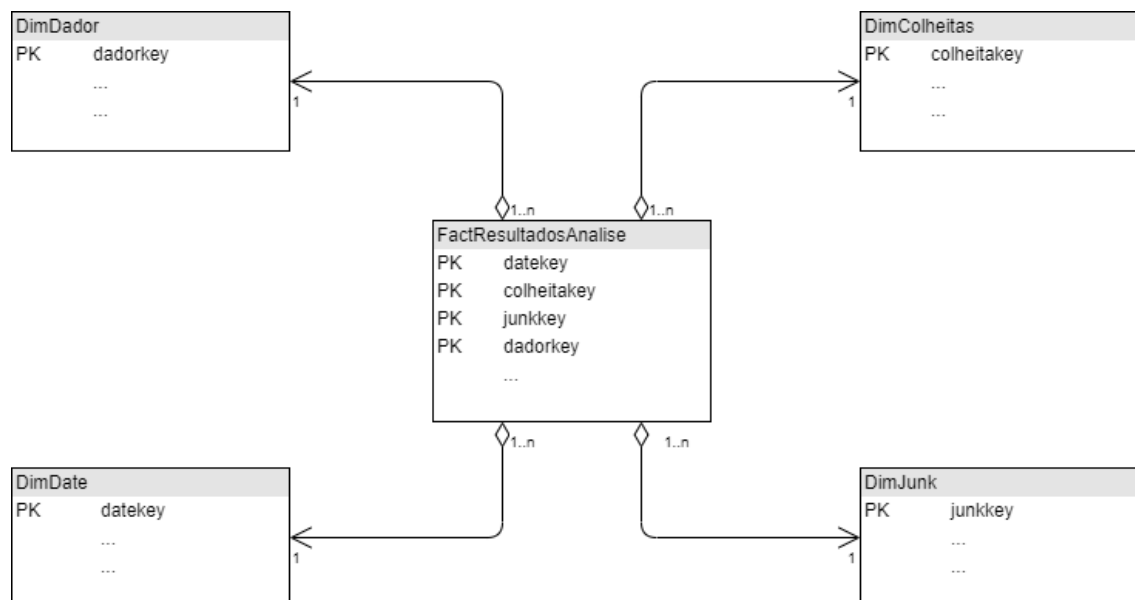


Figura 32 – Representação do Data Mart

6 Implementação da solução

Neste capítulo é descrito todo o trabalho realizado em cada um dos componentes da arquitetura. Numa primeira secção são apresentadas as tecnologias que foram utilizadas na implementação da solução e nas secções seguintes são descritas as etapas do processo de ETL durante o desenvolvimento da solução. Estas etapas refletem as operações de extração que foram necessárias, as operações de transformação aos dados, incluindo as operações de Limpeza e por fim, o carregamento dos dados. Para além disso, são ainda mencionadas as formas de análises de dados a que os profissionais de saúde têm acesso.

6.1 Tecnologias Utilizadas

Para implementar a solução, foram elegidas tecnologias Microsoft, nomeadamente SQL Server para criar o modelo de armazenamento de dados e Microsoft Integration Services para efetuar o processo de integração de dados. Este último, possui características que permitem executar os processos de ETL (Extração, Transformação e Carregamento) e integrar facilmente com o Office e o SQL Server. Para representar a análise aos dados utiliza-se a ferramenta Microsoft Power BI. A escolha das ferramentas é efetuada com base nas características apresentadas no terceiro capítulo e também tendo em conta que o ISEP possui acordo com a Microsoft, para que as ferramentas sejam disponibilizadas gratuitamente.

6.2 Processo de ETL

Em capítulos anteriores já se falou acerca do processo de ETL. Em resumo, este processo consiste na extração de dados das fontes, transformação desses dados tendo em conta as regras de negócio e por fim, carregamento dos dados no repositório de destino, para poderem ser analisados. A Figura 33 representa a arquitetura que caracteriza a solução final. Esta já tinha sido apresentada no capítulo anterior, no entanto é sobre cada uma das etapas que a constituem que se pretende falar nas subsecções seguintes.

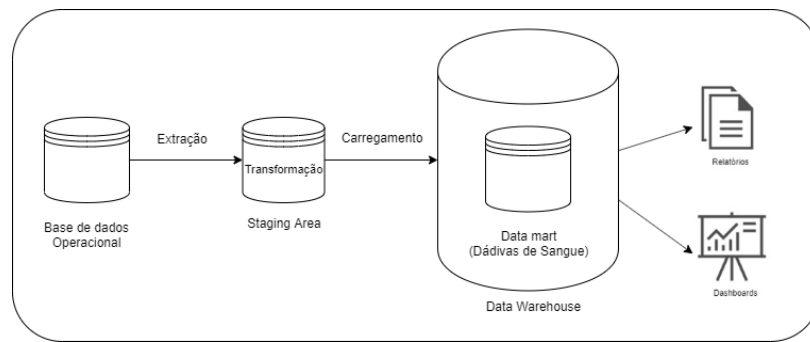


Figura 33 - Processo de ETL na arquitetura do sistema

6.2.1 Operações de Extração

A “Extração” é a primeira fase no processo de ETL. A informação que se encontra na base de dados Access é extraída para uma Staging Area, permanecendo os dados nas tabelas dessa estrutura temporariamente. A Figura 34 representa graficamente o processo de extração na arquitetura proposta para a solução.

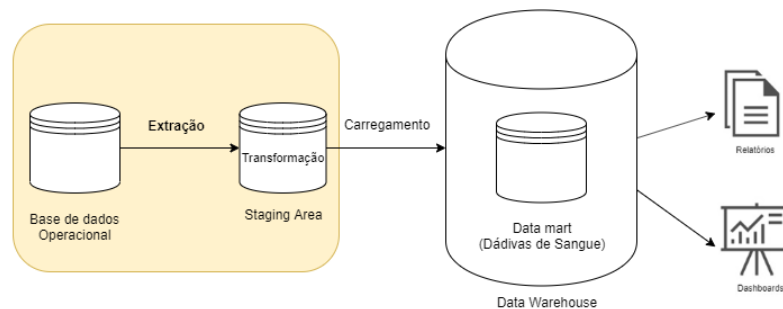


Figura 34 - Fase de extração na arquitetura do sistema

De forma a facilitar o carregamento da informação nas tabelas da Staging Area, as mesmas não possuem qualquer relacionamento, tal como foi possível visualizar na Figura 26 do capítulo anterior. O carregamento dos dados nalgumas das tabelas, como é o caso das “BRIGADAS”, “HOSPITAIS”, “COMPONENTES”, “CONCLUSOES”, “INUTILIZACOES”, “SEXO”, “ESTADO_CIVIL”, “PROFISSOES”, “TIPO DADIVA” foi efetuado de forma direta, enquanto que nas tabelas de “DADORES”, “COLHEITAS” e “ANALISES” foi necessário aplicar regras, proceder-se à remoção de duplicados e até ao envio de dados incoerentes para ficheiros que armazenam os problemas de qualidade dos dados, para que a organização de saúde possa validar e corrigir os mesmos junto dos dadores e mais tarde possam novamente ser carregados para o sistema.

Ainda nesta fase, verificou-se que a Staging Area poder-se-ia resumir às tabelas “DADORES”, “COLHEITAS” e “ANALISES” já que estas possuem no seu conteúdo informação acerca das

restantes tabelas. Posto isto, e de forma a não existir redundância na informação que se pretende disponibilizar, foram efetuadas etapas de transformação, incluindo operações de limpeza nestas últimas tabelas.

6.2.2 Operações de Transformação

Nesta fase é necessário interpretar os dados que se encontram nas tabelas da Staging Area de acordo com as regras de negócio acerca das Dádivas de Sangue e efetuar todas as operações necessárias, quer ao nível de esquema, quer ao nível do valor de dados. A Figura 35 representa a etapa de Transformação na arquitetura proposta para a solução.

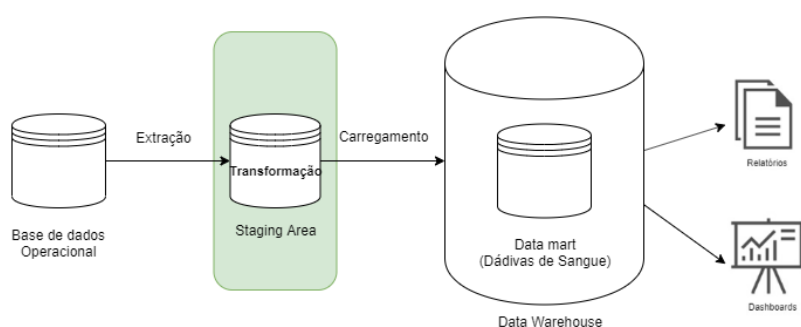


Figura 35 - Fase de transformação na arquitetura do sistema

Nesta etapa, são apresentadas as transformações e operações de limpeza que se realizaram sobre os dados das tabelas “DADORES”, “COLHEITAS” e “ANALISES” na Staging Area, sendo estas apresentadas nas subseções que se seguem.

6.2.2.1 Dadores

A Figura 36 representa o processo de ETL realizado na tabela de “DADORES”. Nesta tabela houve necessidade de realizar algumas transformações e limpeza aos dados, sendo mencionadas tais operações de seguida.

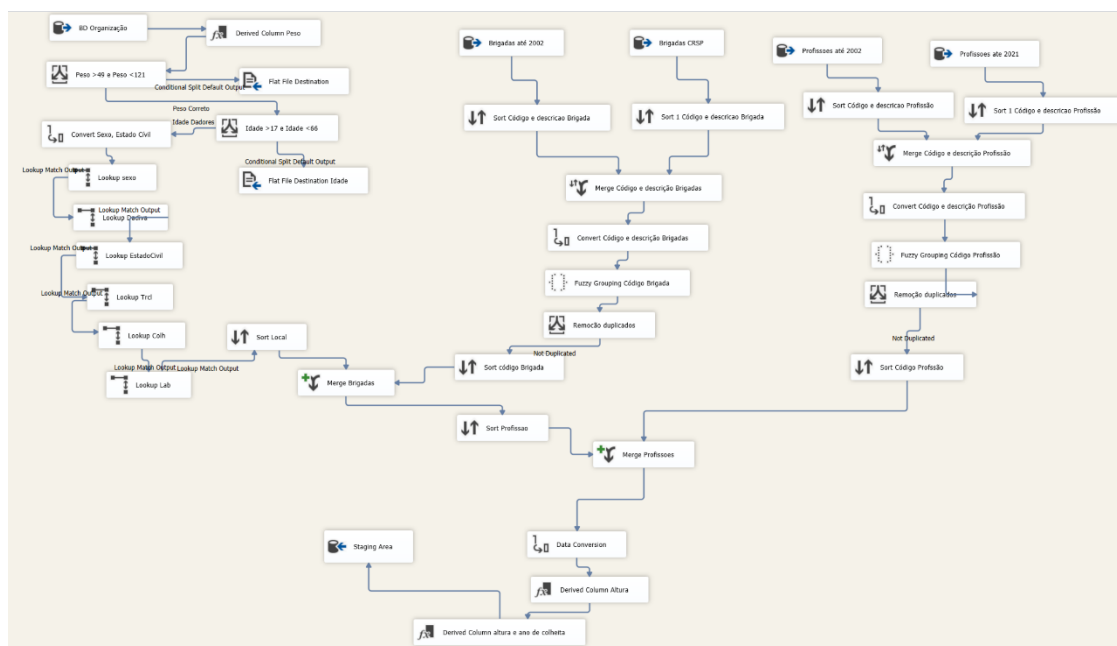


Figura 36 - Processo ETL da tabela "Dadores"

Transformação dos dados

Houve necessidade de realizar algumas conversões relativamente aos tipos de dados, para que os mesmos se aproximassem o mais possível do repositório de destino e evitar que problemas futuros de carregamento possam existir. É exemplo disso, a alteração do tipo de dados para o código e descrição da Brigada e da Profissão, uma vez que na base de dados original estes encontravam-se como Unicode string e foram convertidos para o tipo string.

Para além houve ainda transformações nos tipos de dados para o Sexo e Estado Civil em que na base de dados original os campos eram do tipo double-precision float e passaram no início do processo de ETL a integer e Unicode String, respetivamente.

Limpeza de dados

- **Peso:** Os dadores de sangue têm de ter peso igual ou superior a 50 kg. Considerou-se que todos os registos em que o peso fosse inferior a 50 kg e superior a 120 kg seriam enviados para um ficheiro que contém os problemas de qualidade de dados, para que mais tarde essa informação ser validada pela organização. Para além disso, pesos superiores a 120 kg não refletem indivíduos com hábitos saudáveis, pelo que também vão de encontro às regras para ser dador. Verificou-se ainda, que em relação ao peso, existiam dois registos, onde os valores eram negativos. Desta forma, esses registos foram também enviados para o ficheiro já mencionado e que denomina de "ErrorPeso".
- **Idade:** Para se ser dador é necessário que a idade se encontre compreendida entre os 18 e os 65 anos. Na altura em que os registos foram efetuados verificou-se que existiam alguns casos em que tal não se refletia, pelo que estes registos foram enviados para um

ficheiro de problema de qualidade de dados, denominado de “ErrorIdade”, para mais tarde serem apurados pela organização junto dos dados.

- **Altura:** A informação sobre a altura não influencia para a dívida, no entanto considerou-se que os registos que continham o valor para a altura superior a 2,5 metros, seria necessário efetuar transformação. Quando se realizam transformações nos dados, geralmente é o cliente do produto que toma a responsabilidade de decisão. Foi decidido calcular a média de todos os registos que tinham altura até 2,5 metros e colocar essa média nos registos em que a altura é superior a 2,5 metros.
- **Remoção de duplicados:** A informação acerca das Brigadas e das Profissões encontram-se em várias tabelas. Foi necessário realizado um Merge e após os dados se encontrarem carregados, foi utilizada uma expressão para remover valores duplicados. Assim, o processo de ETL fica preparado para introdução de novas informações, sem que a esta se repeta.

6.2.2.2 Colheitas

A Figura 37 representa o processo de ETL realizado a tabela “COLHEITAS”. Nesta também foi necessário aplicar operações de transformação e limpeza aos dados, sendo estas mencionadas de seguida.

6.2.2.3 Análises

A Figura 38 representa o processo de ETL desenvolvido para a tabela de “ANALISES”. Nesta tabela também foi necessário recorrer a operações de transformação e limpeza aos dados, pelo que de seguida, as mesmas são mencionadas.

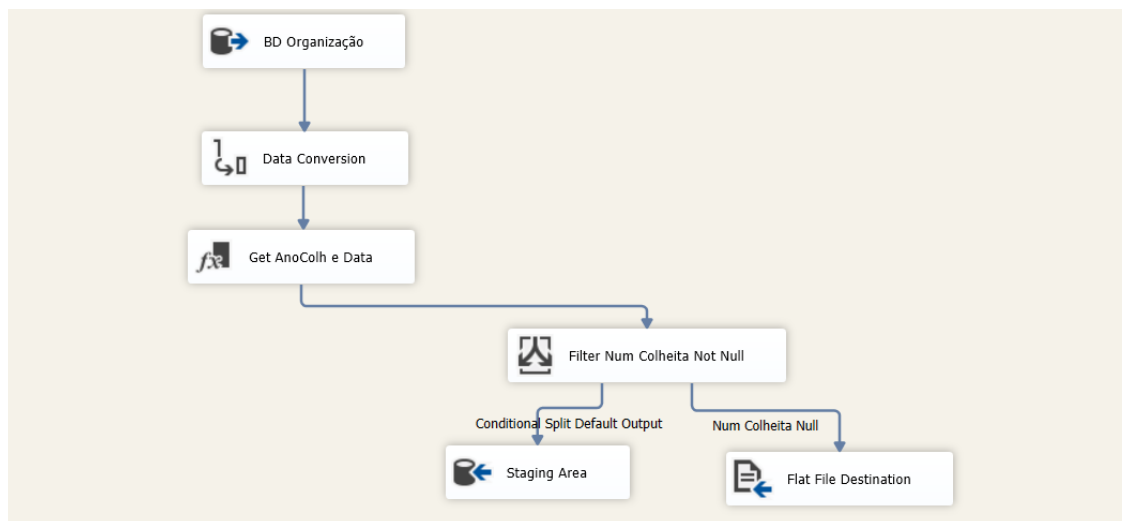


Figura 38 - Processo ETL da tabela "Análises"

Transformação dos dados

Foi necessário efetuar conversões aos tipos de dados de cada uma das colunas referentes aos testes que se iriam realizar nas análises de sangue. Posto isto, todos esses atributos que se encontravam como Unicode string passaram para o tipo string. Foram também realizadas transformações ao formato da data.

Limpeza de dados

- **Número de Colheita:** Nas análises de sangue, para além dos resultados obtidos para cada um dos testes realizados, é essencial ter a informação de data, número de dador e número da colheita. Desta forma, todos os registos que não possuem número de colheita foram enviados para um ficheiro de problema de qualidade de dados, denominado de “ErrorData”, para mais tarde poderem ser analisados e apurados pela organização e posteriormente a informação ser carregada no sistema.

6.2.3 Operações de Carregamento

Após os dados se encontrarem carregados na Staging Area, segue-se o processo de carregamento dos mesmos para o Data Mart. A Figura 39 representa a etapa de Carregamento na arquitetura proposta para a solução.

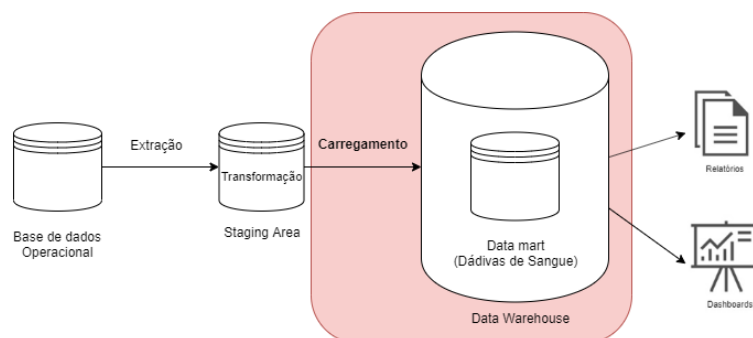


Figura 39 - Fase de carregamento na arquitetura do sistema

Antes de implementar o Data Mart recorreu-se à criação de um Logical Data Map, onde se procede ao mapeamento de cada atributo, da base de dados operacional para a Staging Area e desta última para o Data Mart (Wisdom Jobs, 2019).

O Logical Data Map é apresentado sobre a forma de um ficheiro de Excel e possui os seguintes componentes:

- **Componentes Fonte:** é identificada a base de dados origem, o nome da tabela, nome da coluna e o tipo de dados.
- **Componentes Destino:** é identificada o nome da tabela de destino, o nome da coluna, o tipo de tabela (Dimensão ou Facto), tipo de dados e tipo de SCD.
- **Transformação:** Corresponde ao tipo de conversão de dados para o formato esperado. Pode ser expresso em SQL ou pseudocódigo. Quando o campo não é preenchido, significa que os dados são carregados diretamente para o destino, não necessitando de quaisquer transformações.

O Anexo 1 e 2 representam o Logical Data Map criado para auxiliar na criação do Data Mart. Quando se efetua o carregamento dos dados da Staging Area para o Data Mart, numa primeira fase são carregadas as dimensões e só depois é que se executa o carregamento da tabela de Factos, pois esta última depende da chave de cada uma das dimensões e necessita de ter a informação sempre atualizada.

Quando se está perante o carregamento das dimensões é necessário validar se os atributos que as compõem devem sofrer alterações, quer para conseguir-se manter o histórico dos dados, quer para atualizar a informação (manter sempre a informação mais atual). No Data Mart implementado, os atributos das dimensões apenas sofreram alterações do Tipo 1 e do Tipo 2.

Posto isto, foram inseridos três atributos nos scripts do Data Mart para guardar as datas: “Actual”, “Efectivo”, “Expirado”, com o intuito de perceber de imediato qual a informação mais atual para determinado dador, quanto ao registo, colheita e análise de sangue.

A Figura 40 corresponde ao processo implementado para se efetuar o carregamento da dimensão “DimDadores”. O carregamento das tabelas “DimColheitas” e “DimJunk” foi realizado de forma semelhante.

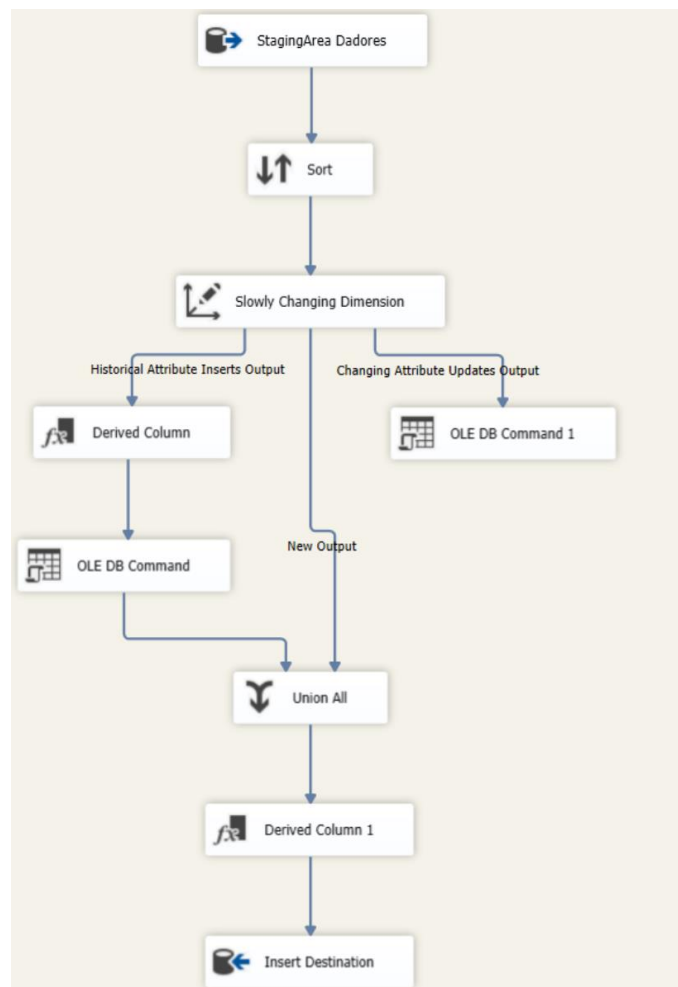


Figura 40 - Carregamento da tabela DimDadores

Após as dimensões estarem carregadas, é necessário proceder-se ao carregamento da tabela de Factos. A Figura 41 corresponde o processo implementado para o carregamento da tabela de Factos.

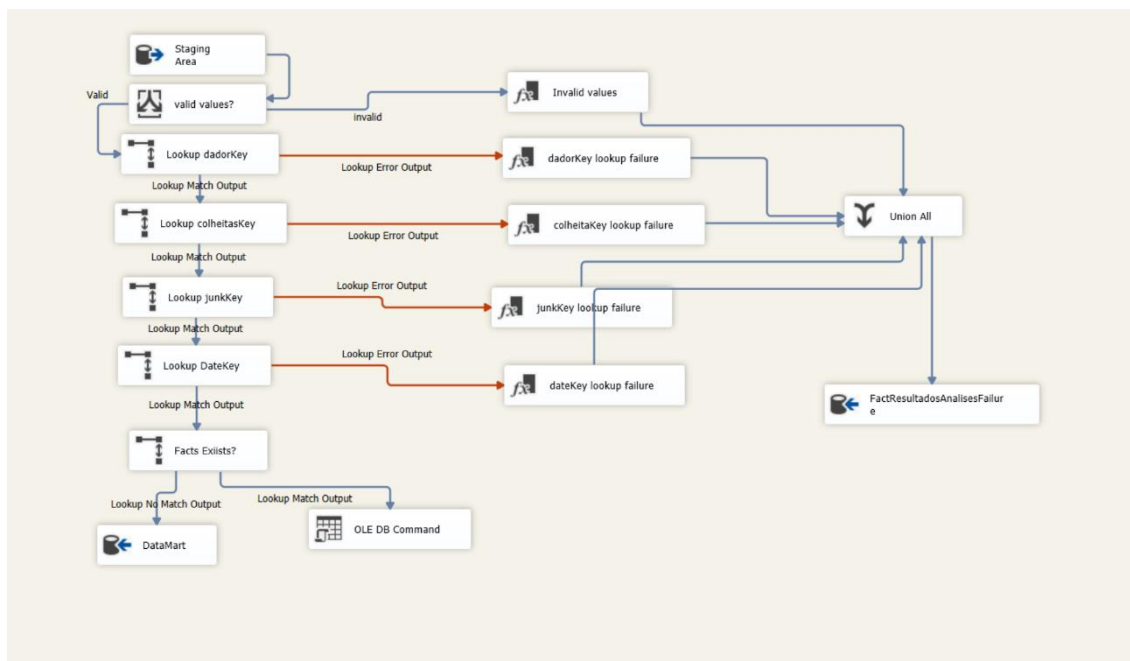


Figura 41 - Carregamento da tabela de factos

Neste processo, a informação é validada, de forma a enviar para a tabela de destino apenas a informação que obedece às regras de negócio, ou seja, só são enviadas as análises que possuem data, número de dador e número da colheita. Foram criadas as chaves, para que esta tabela possa efetuar as ligações com as restantes e por fim, toda a informação numérica relativa aos testes de sangue que se considere essencial para obtenção de conclusões. A informação que não se encontre de acordo com as regras de negócio, são enviadas para uma tabela com problemas de qualidade de dados, denominada “FactResultadosAnalisesFailure”, para que os valores acerca das análises que foram efetuados por um dador numa determinada data e colheita possam ser analisados mais tarde.

Por fim, pode-se visualizar, de uma forma geral, como todos os processos já mencionados neste capítulo então interligados. Para tal, é apresentado no Anexo 3 a vista de todo o Control Flow desenvolvido para esta dissertação. Todos os processos já apresentados são alguns exemplos do trabalho realizado ao nível do Data Flow do SQL Server Integration Services da Microsoft.

6.3 Análises de dados

Esta é a última fase apresentada na arquitetura do sistema, sendo visível na Figura 42 a contextualização da mesma.

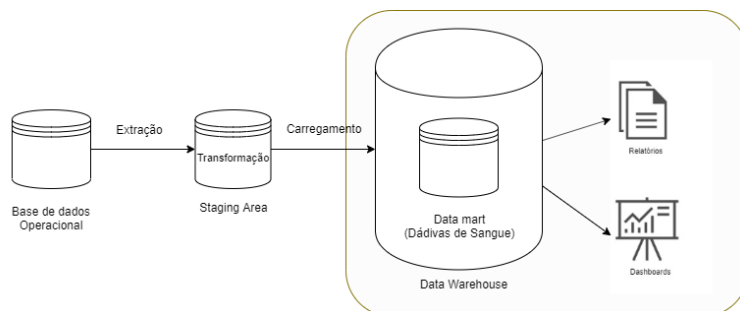


Figura 42 - Fase de análise de dados na arquitetura do sistema

Para se visualizar os dados que se encontram no Data Mart utiliza-se a ferramenta Power BI. Para tal, é necessário efetuar o carregamento de dados do Data Mart para a ferramenta. O Anexo 4 evidencia como é possível efetuar esse procedimento. Pretende-se que os dados sejam apresentados ao utilizador final sobre a forma de dashboards, de forma a que estes consigam de imediato retirar conclusões de acordo com os filtros que aplicam e ainda ser possível executar relatórios, caso prefiram este meio de análise.

No capítulo seguinte são apresentados todos as dashboards que os profissionais de saúde conseguem visualizar. Estas possuem indicadores acerca da informação dos Dadores, das Colheitas e dos resultados às análises de sangue.

7 Apresentação e Análise de Resultados

Neste capítulo são exibidas todas as análises conseguidas com a solução implementada.

As dashboards desenvolvidas, encontram-se de acordo o estudo que foi feito ao longo desta dissertação, evidenciando assim, toda a informação que se considera relevante para o utilizador final.

7.1 Indicadores

Existe um conjunto de indicadores que são alvo de estudo ao longo desta dissertação e que são exibidos ao longo de um conjunto de relatórios. Destacam-se indicadores sobre Dadores de Sangue, sobre Colheitas e sobre o resultado efetuado a Análises de Sangue. O conjunto desses indicadores possibilita rastrear números e motivos que levam a conclusões sobre Dádivas de Sangue.

As análises implementadas têm como utilizador final os profissionais de saúde, pelo que estas devem apresentarem-se visualmente de fácil interpretação, de forma a conseguir-se retirar conclusões rapidamente e com pouco esforço.

7.2 Dashboards

São as dashboards que permitem aos profissionais de saúde monitorizar os resultados relativos às Dádivas de sangue através de painéis, onde visualmente se destacam gráficos e tabelas dinâmicas. Nas subsecções seguintes são evidenciadas todas as dashboards desenvolvidas com informação acerca dos Dadores, das Colheitas e dos resultados obtidos nas Análises de Sangue.

7.2.1 Informação Dadores

A Figura 43 representa o primeiro relatório implementado relativamente à informação de dadores de sangue. Neste painel, os dados podem ser obtidos durante os anos de 2000, 2001 e 2002.

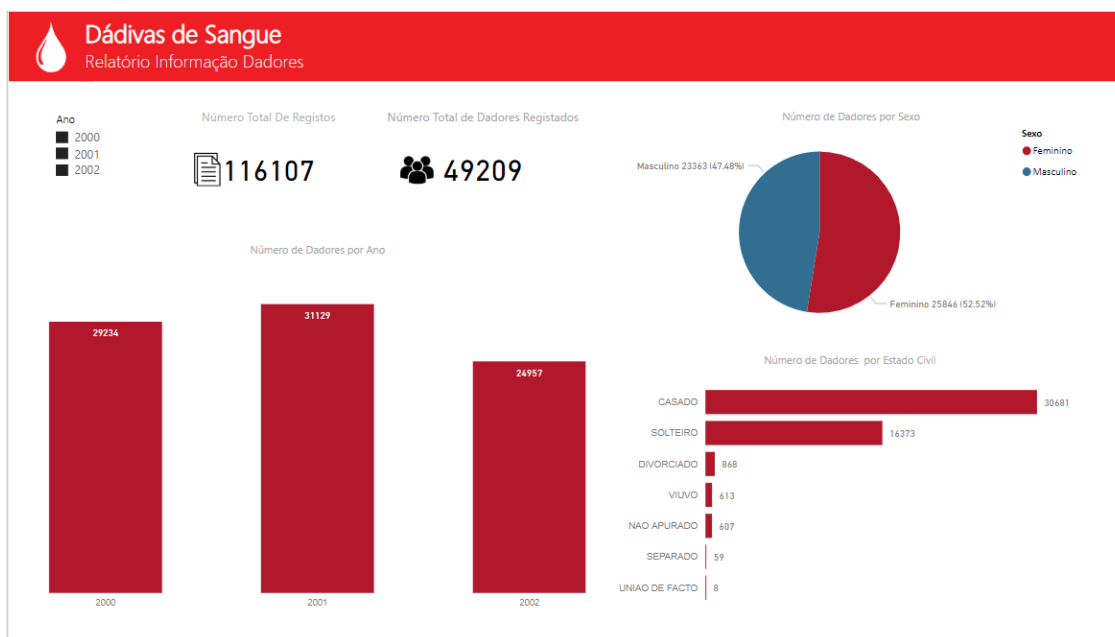


Figura 43 – Relatório informação dos dadores

De acordo com o mesmo, pode-se verificar que durante os três anos foram efetuados 116107 registos (dado os homens poderem doar sangue quatro vezes ao ano e as mulheres três vezes ao ano). Apesar disto, apenas se contabilizam 49209 dadores.

Relativamente ao número de dadores, pode-se visualizar que no ano de 2000 se contabilizaram 29234 dadores, no ano de 2001 houve 31129 dadores e no ano de 2002 houve 24957 dadores. O gráfico relativo ao “Número de Dadores por Ano”, tem a particularidade de se conseguir fazer drill down, para se obter a informação de forma mais detalhada. Através do drill down o valor do número de dadores é conseguido para o quadrimestre, trimestre, mês e até mesmo ao dia correspondente a cada um dos anos.

A Figura 44 representa o drill down que se efetuou no ano de 2001, terceiro quadrimestre, mês julho, dando os valores dos dadores por dia.

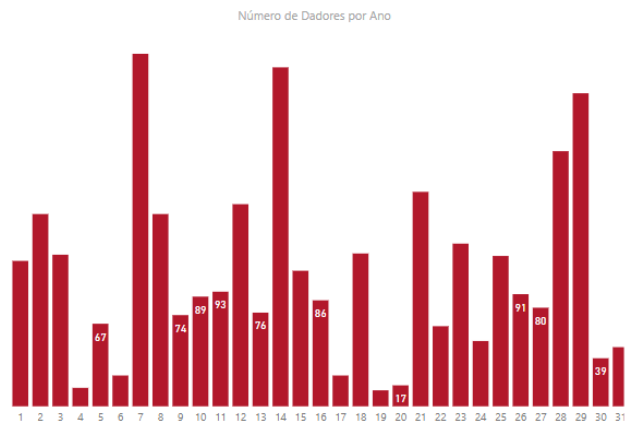


Figura 44 - Drill down do Número de dadores por Ano

Quando se passa com o rato sobre cada uma das barras, surge toda a informação para ser possível contextualizar o número de dadores no tempo e o respetivo valor.

Ainda na dashboard, também se pode visualizar o número de dadores por sexo, sendo que o número relativo ao sexo feminino é de 25846 dadores e relativo ao sexo masculino é de 23363 dadores. Desta forma, verifica-se que o sexo feminino possui uma percentagem de 52,52%, enquanto que o sexo masculino possui uma percentagem de 47,48%.

Por fim, pode-se ainda visualizar o número de dadores por estado civil. Os três estados civis que mais se destacam são: “Casado” com 30681 dadores, “Solteiro” com 16373 dadores e por fim, “Divorciado” com 868 dadores. Por último, os estados civis que possuem menos dadores são “Separado” e “União de Facto” com e 59 e 8 dadores, respetivamente.

A Figura 45 representa um segundo relatório, ainda com informação sobre os dadores de sangue.

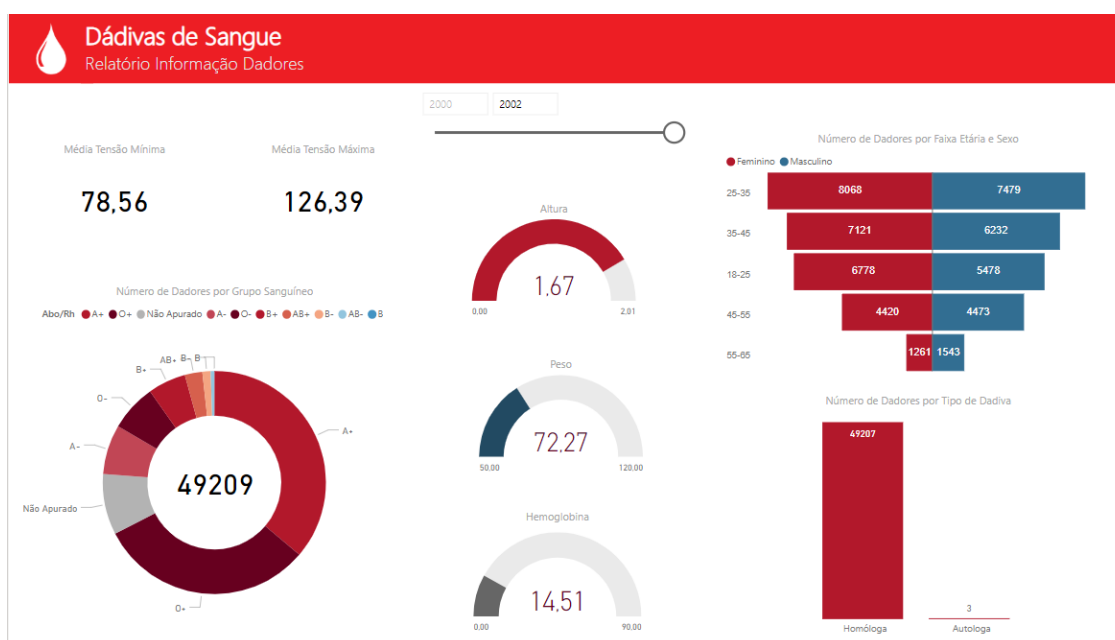


Figura 45 - Relatório informação dos dadores

De acordo com o mesmo, a informação nele contida pode ser visualizada ao longo dos três anos já mencionados.

São apresentados os valores médios para a tensão máxima (sistólica) e mínima (diastólica).

De acordo com a informação obtida, o valor para a média da tensão máxima encontra-se “Normal”, uma vez que 126,39 se encontra compreendido entre 120 a 129. O valor da média para a tensão mínima é considerado ótimo uma vez que é inferior a 80.

Também são evidenciados gráficos que contêm o valor mínimo, médio e máximo para a Altura, Peso e Hemoglobina.

Em geral, os dadores possuem em média 1,67 metros de altura, 72.27 kg de peso e 14.51 g/dl de Hemoglobina. A média para a hemoglobina considera-se que se encontra nos valores normais quando comparados com os valores de referência, já estudados no segundo capítulo desta dissertação.

Para além disso, consegue-se ainda visualizar o número de dadores por grupo sanguíneo. De acordo com o gráfico, o grupo sanguíneo que predomina nos dadores é o A+ com 17179 dadores e o que menos predomina é o B, sendo que nos registos, apenas 1 dador possui este tipo de sangue. Apesar de existir o grupo B+ e B-, existem casos em que o ABO não possui subgrupo, considerando então o Rh null. Estima-se que desde 1961 apenas foram detetados 43 casos com esta situação, considerando estes indivíduos como os chamados “Sangue Dourado” (Costa, 2019).

O RH null deve-se ao facto de os glóbulos vermelhos não possuírem nenhum antigénio do tipo RH, não podendo ser considerado positivo (glóbulos vermelhos possuem antigénios com RH positivo) e não pode ser considerado negativo (glóbulos vermelhos possuem antigénios com RH negativo).

Validar qual a faixa etária que mais efetua dádivas também é um fator importante. A faixa etária que mais predomina é a dos 25 aos 35 anos, seguida da faixa etária dos 35 aos 45 anos.

O gráfico que evidencia a faixa etária tem também a particularidade de mostrar ainda o sexo dos dadores.

Para além disso, também é visível o número de dadores por tipo de dádiva de sangue. Dos 49209 dadores, a grande maioria efetua dádiva Homóloga, correspondendo a um total de 49207 dadores e apenas 3 dadores efetuam dádiva Autóloga. Com estes valores, pode-se concluir que existe um dador que efetuou os dois tipos de dádivas.

Toda esta análise pode ser obtida por sexo. Para tal, basta clicar sobre a legenda referente ao sexo, selecionado o pretendido, de forma a se conseguir obter os dados para esse filtro. É de realçar que quando se clica sobre cada uma das partes dos gráficos, todos os outros interagem (Ex: clicar sobre o grupo sanguíneo A+ para se saber o número de dadores nos três anos nessas condições).

O Anexo 5 representa a interação que pode ocorrer nas dashboards, quando se selecionam os filtros.

A Figura 46 corresponde a outro relatório, ainda sobre os dadores de sangue. Foi criada uma terceira dashboard que contém informação acerca da profissão dos dadores e da localidade dos mesmos.

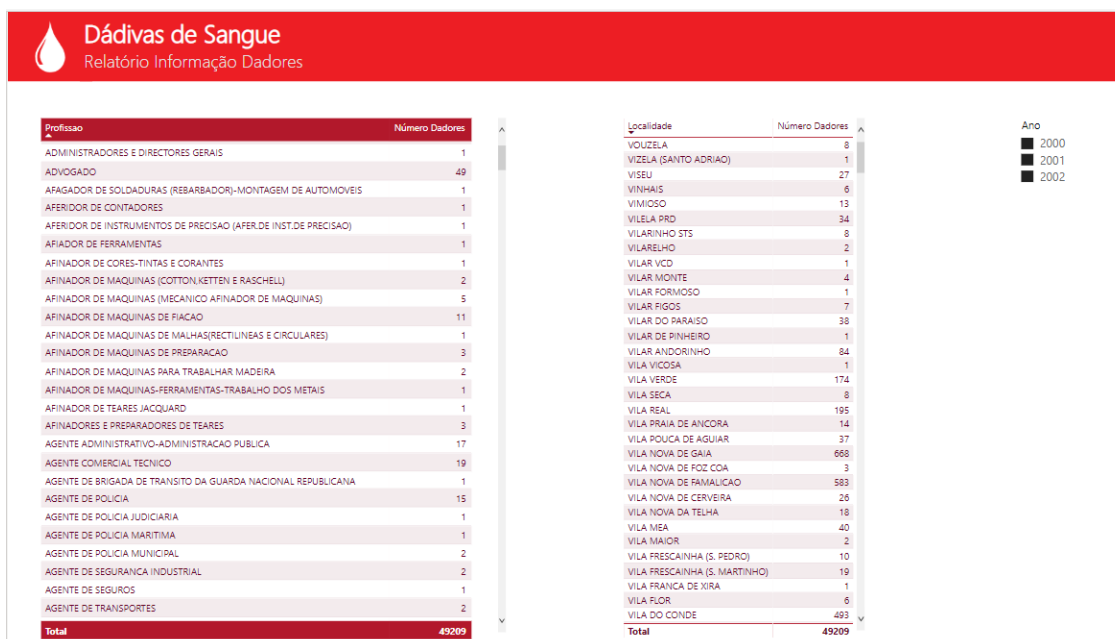


Figura 46 – Relatório informação dos dadores

Também neste relatório é possível obter o número de dadores para cada um destes indicadores ao longo dos três anos. Saber o número de dadores que possui determinada profissão é considerado um fator importante de acordo com o contexto atual. Existem profissões que por sofrerem elevado nível de exposição são consideradas de risco, dado estarem propensas à obtenção de infeções (Ex: trabalhar em enfermarias, matadouros, profissões onde existe exposição a doenças, sangue ou fluidos). Obter tal indicação pode ser útil, já que o profissional de saúde pode identificar circunstâncias que impeçam de imediato a dádiva de sangue (IPST, 2017b).

Este relatório possui o número de dadores pela respetiva profissão e ainda é possível obter o número de dadores pela respetiva localidade. Desta forma, consegue-se saber quais as localidades que possuem maior número de dadores, bem como as profissões dos dadores.

A Figura 47 representa um quarto relatório, ainda sobre os Dadores.

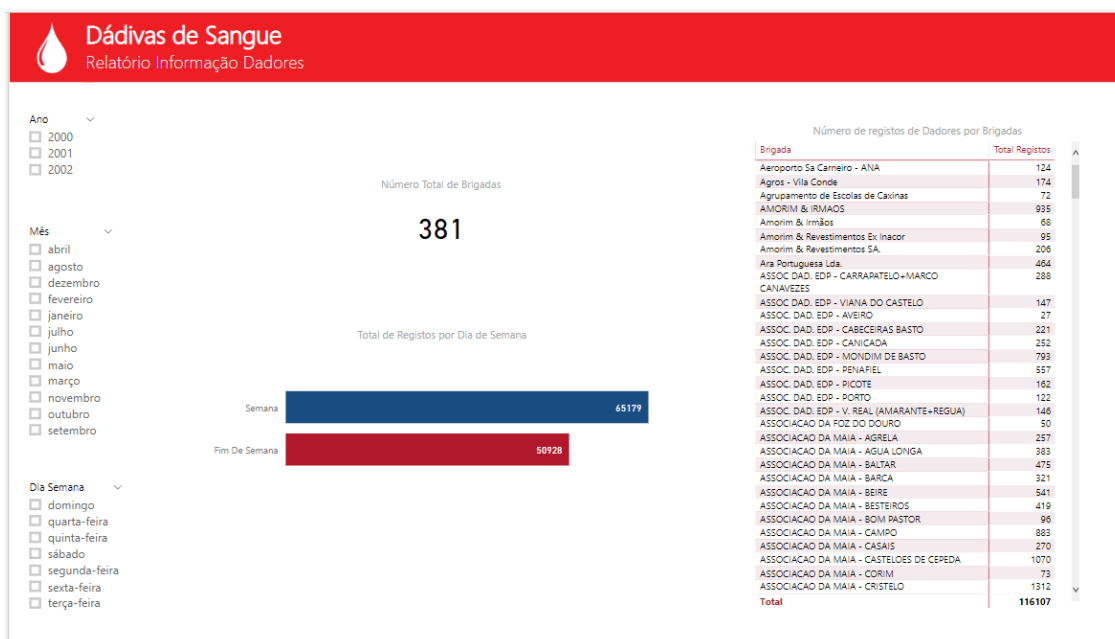


Figura 47 - Relatório de informação dos dadores

Neste relatório, é evidenciado o número de brigadas que efetuaram os registos aos dadores. São as Brigadas que vão ao encontro dos dadores com o objetivo de conseguirem efetuar as colheitas de sangue. Muitas vezes, contam com o apoio de organizadores locais para que estes possam agendar junto das populações o momento mais oportuno para se mobilizarem e levarem os equipamentos, materiais e recursos humanos. As brigadas são promovidas pela organização de saúde que se responsabiliza pela gestão de dádivas de sangue a nível nacional. Nos três anos verifica-se que houve um total de 381 brigadas. É ainda mostrado os nomes das brigadas e número de registos que cada uma efetuou e por fim, salienta-se o número de registos que foram realizados durante a semana ou ao fim de semana.

É de realçar que toda esta informação pode ser obtida ao ano, mês e até mesmo ao nível do dia da semana.

Por fim, a Figura 48, retrata o último relatório acerca dos dadores de sangue.

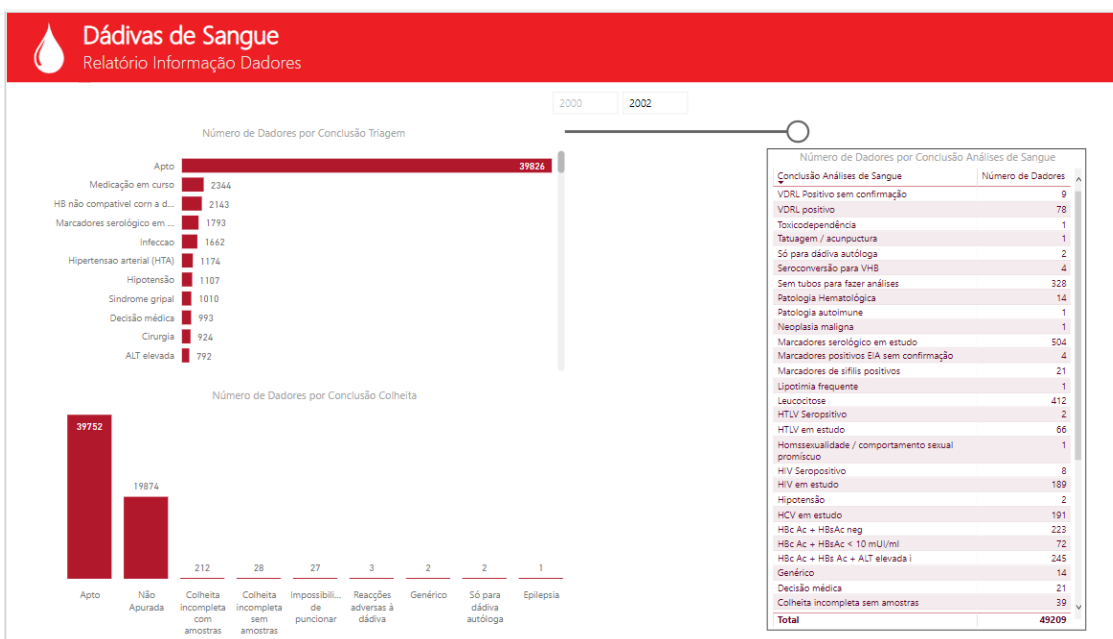


Figura 48 - Relatório de informação dos dadores

Este possui conclusões obtidas quer ao nível da triagem, das colheitas e das análises de sangue, para cada registo, representando o conjunto de todas as conclusões obtidas nas três fases das Dádivas de Sangue. Estas conclusões podem ser obtidas ao longo dos anos de 2000, 2001 e 2002.

Para os profissionais de saúde, quando efetuam para um determinado dador um novo registo, nalgumas situações pode ser útil saber de imediato quais as conclusões obtidas nos registos anteriores.

7.2.2 Informação Colheitas

Acerca das colheitas, pode-se visualizar o primeiro relatório na Figura 49.

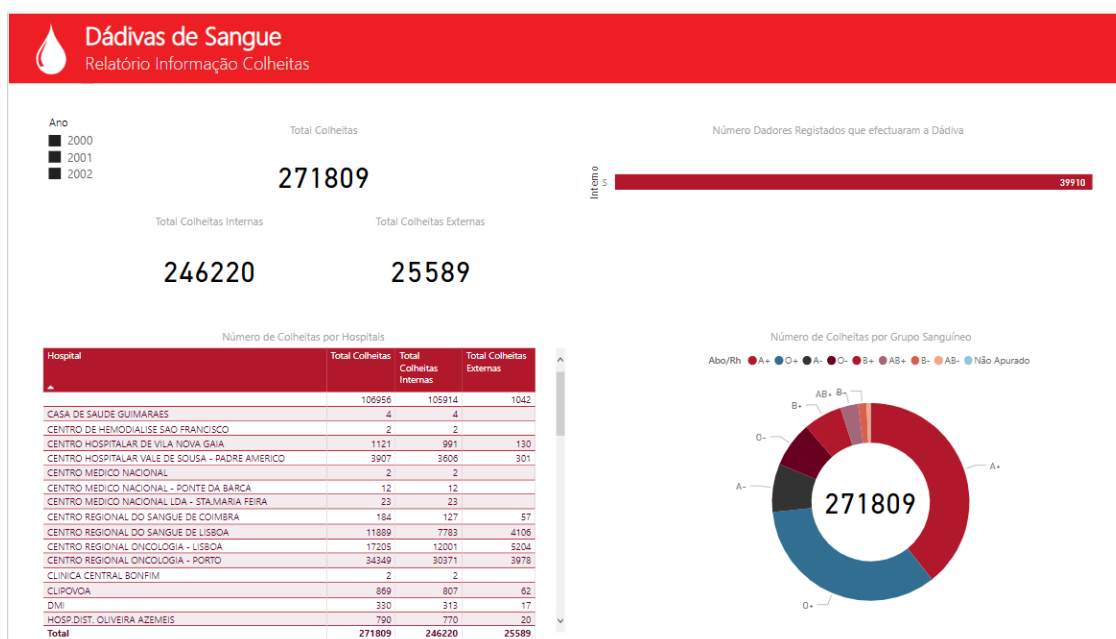


Figura 49 – Relatório informação das colheitas

Neste relatório a informação pode ser obtida durante os três anos (2000, 2001 e 2002) e consegue-se visualizar o número total de colheitas, bem como o número total de colheitas efetuadas por dadores que se encontram registados (“Total colheitas internas”) e o número total de colheitas efetuadas por dadores que não possuem registo (“Total Colheitas Externas”). Estes últimos, não possuem número de dador e por esse motivo não se consegue saber ao certo o número de dadores que se encontram nestas condições.

Durante os três anos, verifica-se um total de 271809 colheitas, das quais 246220 foram efetuadas por dadores registados e 25589 por dadores não registados. Para além disso, é apresentado o número de dadores registados que participaram nas colheitas, sendo estes 39910 dadores.

As colheitas podem ser realizadas por Brigadas e por Hospitais. Ao nível dos Hospitais foi também efetuada uma análise para apurar o número total de colheitas realizadas junto destes e evidenciar quantas colheitas foram realizadas por dadores registados e por dadores não registados. Toda esta informação é visualizada com recurso a uma tabela.

Por fim, é representado o número de colheitas por grupo sanguíneo. Verifica-se que nas colheitas efetuadas o grupo sanguíneo que mais predomina é o A+ com 106827 colheitas e o que menos predomina é o AB- com 1605 colheitas. Também é visível que ainda existem 232 colheitas em que o valor para o grupo sanguíneo não foi apurado.

Foi criado um segundo relatório com a informação das colheitas, que é exibido na Figura 50.

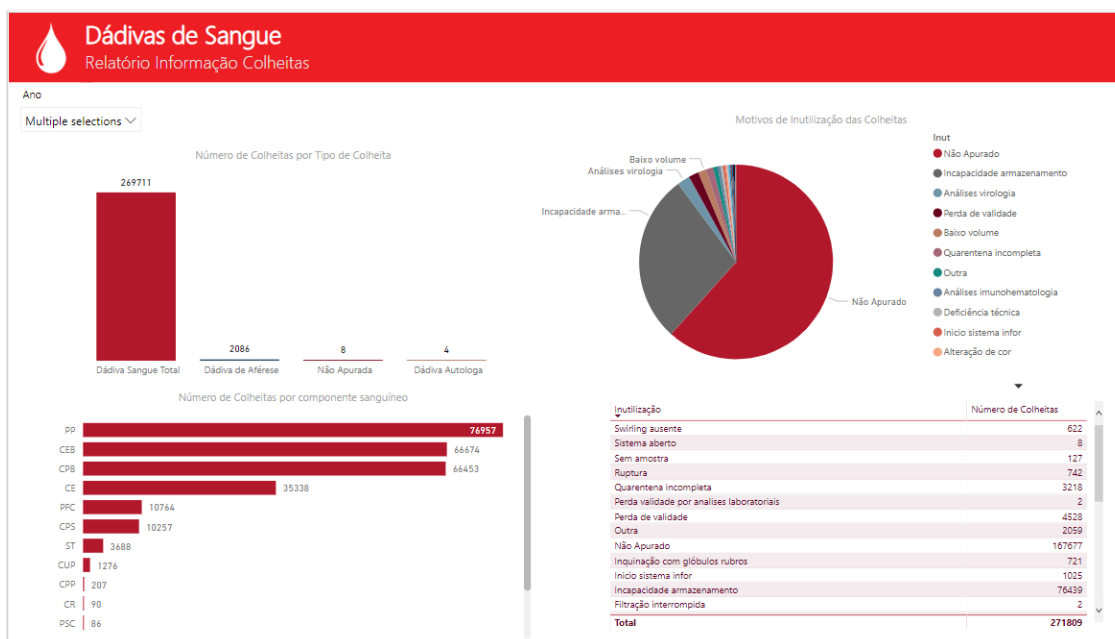


Figura 50 – Relatório informação das colheitas

Neste, consegue-se visualizar a informação para cada um dos anos. Foi apurado o número de colheitas por tipo, verificando-se que 269711 colheitas são dádivas de sangue total, retratando a grande maioria. Para além disso, 2086 colheitas são dádiva Aférese, existem 8 colheitas em que o valor não foi apurado e ainda 4 colheitas em que o tipo é Dádiva Autóloga.

Ainda neste relatório, salientam-se os componentes sanguíneos que foram retirados nas colheitas. Os componentes sanguíneos apresentam-se segundo códigos, uma vez que a descrição para estes não se encontrava na base de dados de origem, ficando esta tarefa para apurar junto da organização, posteriormente. Mesmo assim, verifica-se que os “PP”, os “CEB” e “CPB” são os componentes que foram retirados na maioria das colheitas e os “CP”, “PRP” e “CUG” são os que menos foram retirados.

Por fim, são evidenciados os motivos de inutilização das colheitas. Pode-se visualizar através de um gráfico circular que na maioria o motivo não foi apurado, seguindo-se o motivo “Incapacidade Armazenamento” e “Análise Virologia”. Quando se clica sobre cada uma das partes do gráfico circular, é evidenciado na tabela abaixo o motivo e o respetivo número de colheitas que se encontram nessas circunstâncias.

7.2.3 Informação Resultado Análises de Sangue

Sobre o resultado das Análises de Sangue foram criados duas dashboards, em que a informação pode ser obtida durante o ano, mês, dia de semana e até mesmo ao dia. Para além disso, como também é importante saber se os valores obtidos nos testes realizados às análises de sangue são normais ou não, é importante para alguns deles saber o respetivo valor por sexo, pelo que também foi criado este filtro.

Posto isto, foram obtidos dois relatórios que se encontram representadas na Figura 51 e Figura 52.

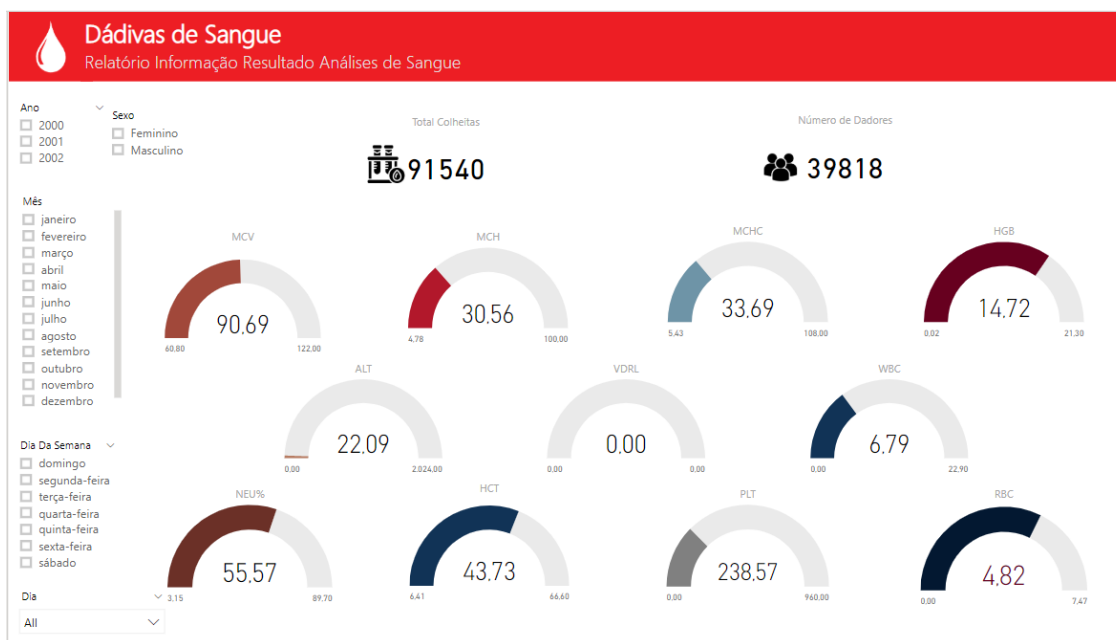


Figura 51 – Relatório informação das análises de sangue



Figura 52 – Relatório informação das análises de sangue

Em ambos os relatórios evidencia-se o número total de colheitas que foram analisadas e o número de dadores que participaram nesta amostra para o estudo em questão.

No primeiro relatório os gráficos representados possuem o valor mínimo, médio e máximo para cada teste já mencionado no capítulo 5, após o diagrama de casos de uso.

De acordo com os gráficos, verifica-se que os valores médios dos testes de ALT, MCV, MCH, MCHC, HGB, NEU%, HCT, RBC se encontram de acordo com os níveis normais estudados no segundo capítulo desta dissertação. Para além disto, verifica-se que para o teste de VDRL não foram apurados valores e por esse motivo o valor médio encontra-se a 0. Mesmo assim, optou-se por colocar o gráfico, uma vez ser um dos testes que aparece na base de dados original, e caso futuramente os valores sejam apurados possam ser refletidos nos relatórios. Para os testes de WBC e PLT os valores médios encontram-se anormais, encontrando-se os mesmos abaixo dos valores normais considerados como referência.

No segundo relatório acerca do resultado de análises de sangue, são ainda representados os valores médios para testes a marcadores serológicos, para detetar a presença do vírus da Hepatite B, C, HIV e sífilis.

De acordo com os valores, dado todos eles serem positivos, pode-se concluir que os valores médios apontam para a presença de infeções. Desta forma, em alguns casos serão necessários novos exames para se poder detetar o material genético do vírus e perceber em que fase da infeção os indivíduos se encontram, pelo que estes indicadores devem ser vistos para cada caso em particular.

7.3 Interface Mobile

Após conclusão de todos os relatórios, foi ainda experimentado a vista “Phone Layout” disponível na ferramenta Power BI Desktop. Nesta vista, os elementos que compõem os relatórios devem ser organizados para serem visualizados em dispositivos móveis.

A Figura 54 representa três exemplos de relatórios com informação de dadores, colheitas e análises de sangue, respetivamente, que foram publicados e acedidos através de um dispositivo móvel.

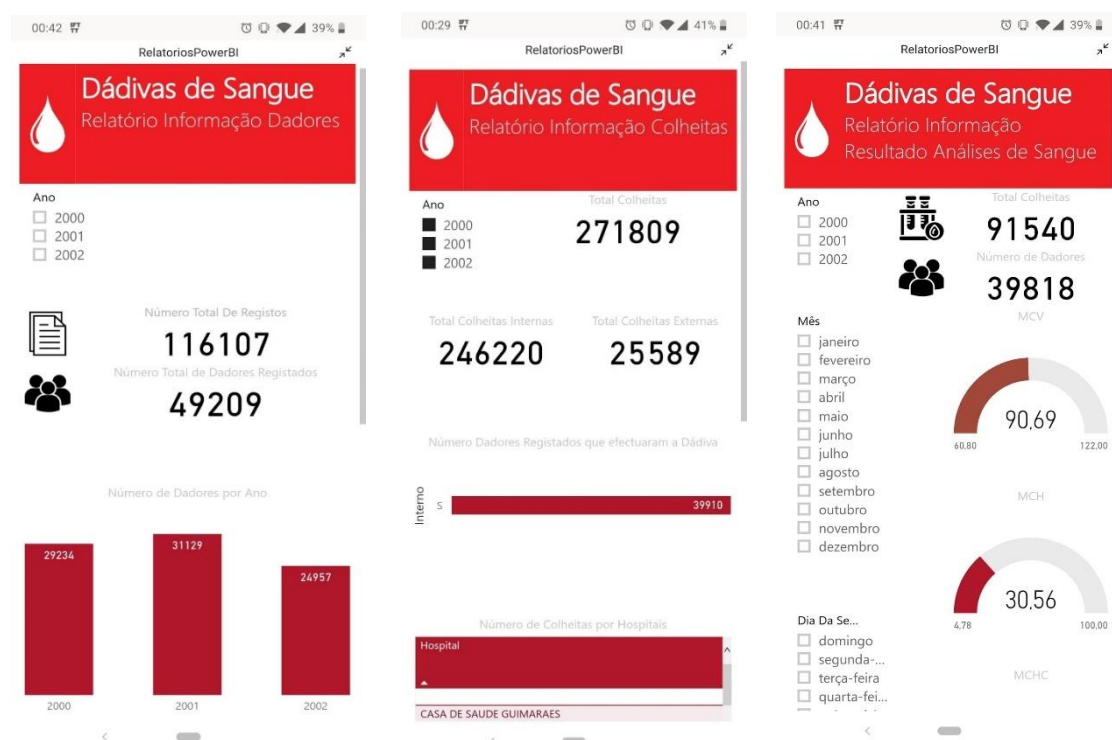


Figura 53 – Modo visualização mobile

8 Avaliação da Solução

Este capítulo tem o objetivo de avaliar o sistema de Business Intelligence que se dedica à análise de dados de Dádivas de Sangue. Para tal, recorre-se a um conjunto de grandezas para avaliar a viabilidade da solução implementada. A solução foi avaliada através de inquéritos de satisfação, onde requisitos como a usabilidade e desempenho são analisados.

8.1 Metodologia de Avaliação

A solução desenvolvida, foi avaliada com recurso a dois inquéritos de satisfação, realizados a dois grupos de profissionais distintos: Profissionais de IT e Profissionais de Saúde.

O objetivo é saber a opinião técnica dos profissionais de IT sobre alguns pontos da aplicação e a opinião dos profissionais de saúde acerca da forma como os dados lhes são apresentados. Desta forma, os inquéritos de satisfação visam retratar pontos de vista diferentes em relação à aplicação.

O grupo de profissionais de IT que responderam ao inquérito, são reflexo de vários anos de experiência em Desenvolvimento de Software e já tiveram oportunidade de trabalhar na área da Saúde. Desta forma, estes conseguem avaliar o sistema e perceber se foi ou não bem conseguido conhecendo a área de negócio e as necessidades dos Sistemas de Saúde.

O grupo de profissionais de saúde que responderam ao inquérito, trabalham no serviço Nacional de Saúde, e com os quais houve oportunidade de falar acerca do trabalho desenvolvido neste último ano. A estes profissionais foram apresentados alguns relatórios que demonstram como a informação acerca das Dádivas de sangue é exibida. Apesar desta dissertação tratar-se de um trabalho académico, torna-se importante a opinião dos profissionais de saúde já que estes são os utilizadores aos quais a aplicação se destina, caso a mesma fosse disponibilizada em ambiente real. Melhor do que ninguém, os profissionais de saúde conseguem ter uma maior capacidade analítica acerca da utilidade deste tipo de aplicação, dado as circunstâncias que enfrentam no dia a dia.

Ambos os inquéritos foram disponibilizados e realizados online, dado ser uma forma rápida de fazer chegar os mesmos aos profissionais e também se torna intuitiva a forma como a informação é apresentada. Para além disso, é uma forma de aproveitar o potencial de ferramentas online que permitem de imediato obter a avaliação acerca das respostas obtidas.

Em ambos os inquéritos de satisfação, foi utilizado um conjunto de afirmações as quais foram classificadas, segundo a escala Likert, para especificar o nível de concordância com as mesmas.

A escala de Likert mede a satisfação do utilizador segundo níveis de classificação. Na solução implementada foram aplicados cinco níveis de satisfação, de forma a ser possível medir o comportamento dos utilizadores relativamente à concordância com as afirmações, a avaliação

do serviço prestado e a frequência de utilização. Nos inquéritos de satisfação foi utilizada a escala que se encontra na Tabela 19.

Tabela 19 - Escala de Likert

| Escala | Descrição |
|--------|----------------------------|
| 1 | Discordo totalmente |
| 2 | Discordo |
| 3 | Não concordo, Nem discordo |
| 4 | Concordo |
| 5 | Concordo totalmente |

O inquérito de satisfação realizado aos Profissionais de IT pode visualizar-se no Anexo 6, enquanto que o inquérito realizado aos Profissionais de Saúde, acerca da avaliação da solução é apresentado no Anexo 7.

8.2 Resultado dos Inquéritos realizados aos Profissionais de IT

O inquérito de satisfação realizado ao grupo de profissionais de IT contou com a participação de 13 pessoas, as quais tinham de responder a um conjunto de quatro questões. De acordo com as opiniões obtidas, em geral, conseguiu-se resultados bastante positivos, os quais podem ser visualizados graficamente no Anexo 6.

Apesar disso, encontra-se de forma resumida a Tabela 20, o resultado para a frequência das respostas a cada questão.

Tabela 20 - Frequência de respostas dos profissionais de IT

| | 1 | 2 | 3 | 4 | 5 |
|----|--------|--------|-----------|-----------|-----------|
| R1 | 0% (0) | 0% (0) | 7,7% (1) | 46,2% (6) | 46,2% (6) |
| R2 | 0% (0) | 0% (0) | 15,4% (2) | 53,8% (7) | 30,8% (4) |
| R3 | 0% (0) | 0% (0) | 15,4% (2) | 46,2% (6) | 38,5% (5) |
| R4 | 0% (0) | 0% (0) | 15,4% (2) | 30,8% (4) | 53,8% (7) |

Segundo os resultados obtidos, na primeira questão houve 6 participantes que selecionaram o nível de concordância 4 (concordo) e outros 6 que selecionaram 5 (concordo totalmente). Para além disso, houve ainda 1 participante que selecionou 3 (Não Concordo Nem Discordo). Posto isto, a maioria está de acordo de que a arquitetura do sistema é adequada ao negócio.

Na segunda questão, 7 participantes responderam 4 (Concordo), refletindo a opinião da maioria, 2 participantes responderam 3 (Nem Concordo Nem Discordo) e ainda houve 4 participantes que responderam 5 (Concordo Totalmente). Também neste ponto se verifica, que a maioria concorda que a solução se encontra preparada para necessidades futuras, nomeadamente para mostrar novos indicadores.

Na terceira questão, 6 participantes mostraram concordância, pelo que responderam 4 (concordo) e 5 responderam 5 (concordo totalmente). Por fim, ainda houve 2 participantes que responderam 3 (Nem concordo nem Discordo). Com isto, também a maioria está de acordo que o sistema está construído de forma a integrar facilmente com outras aplicações.

Por último, e quarta questão, 7 participantes responderam 5 (Concordo Totalmente), refletindo assim a maioria. Ainda se verificou a opinião de 4 participantes com nível 4 (concordo) e 2 com nível 3 (Nem Concordo Nem Discordo). A última questão evidencia que também o tempo de resposta da aplicação é adequado.

8.3 Resultado dos Inquéritos realizados aos Profissionais de Saúde

O inquérito de satisfação realizado ao grupo dos profissionais de saúde contou com o contributo de 15 participantes e este grupo de inquiridos também se mostrou satisfeito com o resultado da aplicação.

O Anexo 7 contém o resultado obtido graficamente para cada uma das questões efetuadas. Para além disso a Tabela 21 representa, de forma resumida, a frequência de resposta a cada uma das 6 afirmações colocadas no inquérito.

Tabela 21 - Frequência das respostas dos profissionais de saúde

| | 1 | 2 | 3 | 4 | 5 |
|----|--------|--------|-----------|-----------|-----------|
| R1 | 0% (0) | 0% (0) | 6,7% (1) | 40% (6) | 53,3% (8) |
| R2 | 0% (0) | 0% (0) | 13,3% (2) | 60% (9) | 26,7% (4) |
| R3 | 0% (0) | 0% (0) | 13,3% (2) | 33,3% (5) | 53,3% (8) |
| R4 | 0% (0) | 0% (0) | 13,3% (2) | 33,3% (5) | 53,3% (8) |
| R5 | 0% (0) | 0% (0) | 13,3% (2) | 60% (9) | 26,7% (4) |
| R6 | 0% (0) | 0% (0) | 6,7% (1) | 53,3% (8) | 40% (6) |

De acordo com a análise dos resultados, na primeira questão, 8 participantes classificaram a afirmação como 5 (concordo totalmente), enquanto 6 participantes escolheram o nível 4 (concordo) e ainda 1 escolheu o nível 3 (Nem concordo Nem Discordo). Desta forma conclui-se que a maioria considerou que a aplicação seria útil e intuitiva.

Na segunda questão, 9 participantes escolheram 4 (Concordo), 2 participantes escolheram 3 (Nem Concordo Nem Discordo) e por fim, 4 escolheram o nível 5 (Concordo Totalmente). Desta forma, mais de metade dos inquiridos concordaram que os filtros se adequam ao tema em estudo.

Na terceira e quarta questão, 2 participantes selecionaram o nível 3 (Nem Concordo Nem Discordo), 5 escolheram o nível 4 (Concordo) e os restantes 8 escolheram nível 5 (Concordo Totalmente). Assim, e mais uma vez, os inquiridos, concordam totalmente que a forma como a informação foi disponibilizada é uma mais valia, torna-se de fácil leitura e análise e é preferível quando comparada a relatórios estáticos.

Na quinta questão, 9 participantes selecionaram o nível 4 (Concordo), 2 participantes selecionaram o nível 3 (Nem Concordo nem Discordo) e ainda houve 4 que selecionaram o nível 5 (Concordo Totalmente). Desta forma, mais de metade dos inquiridos são da opinião de que a aplicação constitui uma boa ferramenta para exploração de dados.

Na última, e sexta questão, 1 dos participantes selecionou o nível 3, 8 selecionaram o nível 4 e ainda 6 selecionaram o nível 5. Assim, segundo os inquiridos a aplicação desenvolvida cumpriu com os objetivos a atingir.

8.4 Conclusão

Os inquéritos de satisfação permitem avaliar a solução segundo dois grupos diferentes de profissionais.

Um dos grupos, os profissionais de IT, avaliaram a solução de acordo com questões mais técnicas enquanto que o outro grupo, os profissionais de saúde, avaliaram a solução na perspetiva de utilizador final.

Com base nas respostas obtidas, e ainda que a amostra tenha sido bastante pequena, conclui-se que o grau de satisfação foi elevado, refletindo-se através de resultados bastante positivos.

Os resultados, evidenciaram que a solução tem interesse e poderia ser útil, caso fosse disponibilizada em ambiente real.

Ainda assim, os inquéritos de satisfação deveriam ter sido disponibilizados a um maior número de pessoas, no entanto, considera-se que o facto de o sistema não ter sido disponibilizado num ambiente real é uma das causas para tal.

9 Conclusão

Neste capítulo são descritas todas as conclusões obtidas ao longo desta dissertação. Para além dos objetivos alcançados, são mencionados os problemas/limitações que existiram ao longo da mesma e por fim, é apresentado o trabalho futuro que ainda se poderia desenvolver, tendo por base o contexto desta dissertação.

9.1 Objetivos Alcançados

Esta dissertação teve como objetivo o desenvolvimento de um sistema de BI que permitisse a gestão de informação acerca de Dádivas de Sangue e que fosse capaz de fornecer toda a informação necessária aos profissionais de saúde, de forma a apoiar no processo de tomada de decisão relativamente ao tema em questão.

O caso de estudo, partir de uma instituição de saúde que se responsabiliza por toda a gestão e monitorização de informação acerca das Dádivas de Sangue e que tem grande importância a nível Nacional.

Ao longo desta dissertação, foi contextualizado o problema, apresentado um estudo acerca dos componentes sanguíneos, apresentou-se a análise de valor, foi ainda analisado o contributo de sistemas de informação na área da saúde, abordados conceitos e explorados a aplicabilidade de sistemas de BI, tendo em consideração o tema em estudo.

Para além disso, foram analisadas um conjunto de ferramentas de ETL e de BI, de forma a verificar quais as que se adequavam melhor às necessidades pretendidas. As ferramentas selecionadas foram o SQL Server Integration Services (SSIS) e o Power BI da Microsoft. Estas duas ferramentas aliadas permitem que o sistema tenha uma melhor performance, exista facilidade de integração e o resultado final torna-se interativo com o utilizador.

O sistema de BI possui uma arquitetura centralizada, de forma a estar preparado para situações futuras que possam advir (Ex: outras fontes de dados diferentes de Access), disponibilizando toda a informação segundo um único padrão e capaz de armazenar histórico.

De forma a melhorar a qualidade do serviço prestado aos profissionais de saúde relativamente às Dádivas de Sangue, foi ainda garantida a filtragem e a execução de relatórios acerca de doadores de sangue, de colheitas realizadas pelos mesmos e de resultado das análises de sangue.

Através dos relatórios implementados e a forma como estes são apresentados consegue-se obter conclusões de imediato, que caso contrário seria impossível. Os relatórios que foram disponibilizados, possuem dados dos doadores de acordo com lei de proteção de dados (myDataPrivacy, 2019), pois os dados referente a estes encontram-se anonimizados, sendo apenas carregados os atributos que não comprometem o utente (Ex: data de nascimento, sexo).

Por fim, como forma de validar se os requisitos funcionais mencionados nesta dissertação foram cumpridos, foram realizados inquéritos de satisfação a uma amostra de profissionais de IT e profissionais de saúde.

Os formulários disponibilizados visam obter a opinião relativamente à usabilidade do sistema, ao grau de satisfação relativo ao tempo de resposta, ao cumprimento dos requisitos e à importância dos indicadores implementados.

Através dos inquéritos de satisfação efetuados aos profissionais de IT e aos profissionais de saúde, ambos os grupos mostram opiniões bastante positivas, que refletem que os objetivos a atingir com a aplicação foram conseguidos.

9.2 Problemas/Limitações

Um dos problemas que se verificou logo à partida nesta dissertação são os dados. A informação disponibilizada não era atual, pelo que o sistema teria mais interesse caso o Data Mart pudesse estar ligado diretamente ao sistema Operacional da organização, de forma a se poder obter a informação em tempo real e não só num determinado período de tempo.

Para além disso, como a dissertação é em âmbito académico, houve dificuldade em entender algumas informações ao nível das regras de negócio sobre as Dádivas de Sangue, e por esse motivo foi desenvolvido um capítulo que servisse de auxílio. Para além disso, houve dificuldade em encontrar profissionais de saúde que pudessem analisar a aplicação com base nas dashboards desenvolvidas e responder aos questionários, uma vez que apesar de serem da área, estes profissionais são externos à organização.

9.3 Trabalho Futuro

Os sistemas de informação contribuem com melhorias para solucionar problemas que têm grande impacto no quotidiano das pessoas.

Na área da saúde, estes são fundamentais porque procuram dar resposta imediata a problemas que até então requerem grandes esforços para solucionar.

Nesta dissertação ainda existe um conjunto de trabalhos que podem ser realizados, para completar o trabalho até então desenvolvido. Destaca-se a adição de novos indicadores relativamente ao tema em estudo, “Dádivas de Sangue”, de forma a justificar o porquê de existem dadores que não pretendem ficar registados no sistema, os motivos que levam os Hospitais a recusar as dádivas e a devolver para outros, perceber as quantidades de sangue que são utilizadas diariamente para tratar doentes e quais as alturas do ano em que são necessárias mais reservas e os respetivos motivos.

Dado que a instituição em estudo atualmente tem as suas funções um pouco alteradas, para além da gestão das dádivas de sangue ainda se responsabiliza pela gestão de órgãos, tecidos e

células de origem humana para transplantação, ficaria também para trabalho futuro, adaptar o sistema para este novo módulo, de forma a ser possível criar um novo data Mart com esta informação e desenvolver novos indicadores relativamente a esta área.

Referências

- 1keydata (2019) *MOLAP, ROLAP, And HOLAP*, 1keydata.
- A.Koen, P. and M.Ajamian, G. (2002) *The PDMA ToolBook for New Product Development*.
- Araújo, E. (2007) *Um estudo sobre as ferramentas OLAP*, DevMedia.
- BioCampello (2019) *Interpretação de Análises Clínicas*.
- Brito, E. (2014) *Mais Sobre DW – Data Marts e Staging Área*, Consultor em TI.
- Brito, E. (2015) *Modelagem Dimensional*, Consultor em TI.
- C-Studio (2017) *“Business intelligence” e analítica: evolução ou complementaridade*, *Jornal de Negócios*.
- Cain, J. (2019) *What is customer value?*, Quora.
- Câmara, B. (2015) *Marcadores sorológicos da Hepatite B*, *Biomedicina Padrão*.
- Costa, C. A. (2019) *RH-null: o que é o ‘sangue dourado’ que menos de 50 pessoas têm?*, *Sábado*.
- Damanpour, F. (2017) *Organizational Innovation*, *Oxford Research Encyclopedias*.
- Deco Proteste (2012) *‘Exames de Diagnóstico Médico’*, 2.
- Dextro, R. B. (2019) *Eritrograma*.
- Diário da República (2007) *‘Decreto-Lei n.º 267/2007, Série I de 2007-07-24’*, pp. 1–21.
- Domo (2019) *Domo, Domo*.
- Economias (2017) *Análise SWOT: o que é e para que serve?*, *Economias*.
- Elias, D. (2014) *O que significa OLTP e OLAP na prática?*, *Canal Tech*.
- Elias, D. (2018) *Arquitetura do Business Intelligence, BI na Prática*.
- Eriksson, U. (2015) *The difference between functional and non-functional requirements*.
- G2 Crowd (2019) *Best Embedded Business Intelligence Software*, *G2 Crowd*.
- Gabry, M. (2016) *O que é proposta de valor e por que é tão importante para seu negócio?*, *Administradores*.
- Gartner (2018a) *‘Magic Quadrant for Business Intelligence and Analytics Platforms’*.

- Gartner (2018b) *Magic Quadrant for Data Integration Tools*, Gartner.
- George, S. (2012) *Inmon or Kimball: Which approach is suitable for your data warehouse?*, ComputerWeekly.com.
- Geovana, N. (2015) *Para que serve o exame HCV?*, Médico Responde.
- GS1 (2017) *Qualidade de dados: o que é e qual a importância para negócios?*
- Guru99 (2019) *What is Dimensional Model?*, Guru99.
- Hemominas (2014) *Componentes e tipos sanguíneos*.
- HIV.gov (2017a) *How Is HIV Transmitted?*, HIV.gov.
- HIV.gov (2017b) *What Are HIV and AIDS?*, HIV.gov.
- IBM (2019) *IBM Cognos Analytics*, IBM.
- IPST, I. (2017a) *Quais os meus direitos enquanto doador de sangue?*
- IPST, I. (2017b) *Será a minha profissão de risco? Poderei dar sangue?*
- IPST, I. (2018) *Sou hipertenso/a. Não me deixam dar sangue se tiver a tensão elevada?*
- Keep.I (2018) *Entenda a Diferença entre Dashboard e Relatório*.
- Lapa, J., Bernardino, J. and Figueiredo, A. (2014) *A comparative analysis of open source business intelligence platforms*.
- Leiderfarb, L. (2018) *Doar bem sem olhar a quem*, Expresso.
- Lemos, M. (2018a) *Leucócitos: O que são e Valores de referência*, Tua Saúde.
- Lemos, M. (2018b) *O que os valores alterados da hemoglobina podem indicar*, Tua Saúde.
- Lemos, M. (2018c) *O que significa Plaquetas alta ou baixas e como identificar*, Tua Saúde.
- Lemos, M. (2018d) *Para que serve o exame ALT Alanina Aminotransferase*, Tua Saúde.
- Lemos, M. (2018e) *Tipos de sangue: para quais pode doar e como doar*, Tua Saúde.
- Lemos, M. (2019) *O que é o hematócrito (Hct) e o que significa quando está alto ou baixo*, Tua Saúde.
- Looker (2019) *Model your Data with LookML.*, Looker.
- Magalhães, L. (2018) *Sangue: função, componentes e tipos*, Toda Matéria.
- Marins, C. S., Souza, D. de O. and Barros, M. da S. (2019) 'O USO DO MÉTODO DE ANÁLISE HIERÁRQUICA (AHP) NA TOMADA DE DECISÕES GERENCIAIS – UM ESTUDO DE CASO'.

Marques, J. R. (2017) *ENTENDA O CONCEITO DE VALOR PERCEBIDO PELO CLIENTE*, Portal IBC.

MediLab (2017) *Conheça as vantagens do BI na área da saúde*, MediLab.

Microsoft (2012) *OLAP Cubes in the SCSM Data Warehouse : Key Concepts*, TechNet.

Microsoft (2018) *SQL Server Integration Services*, Microsoft.

Microsoft (2019a) *Noções básicas do DAX no Power BI Desktop*.

Microsoft (2019b) *O que é o Power BI Desktop?*

Microsoft (2019c) *Power BI*, Microsoft.

Mindtools (2019) *Porter's Value Chain*, Mindtools.

myDataPrivacy (2019) *RGPD | Regime Geral de Proteção de Dados*.

NuWave Solutions (2018) *Data Warehouse Design Techniques – Slowly Changing Dimensions*.

Oliveira, D. K. de and Alves, D. R. (2012) 'Business Intelligence aplicado a área da saúde: potencializando a tomada de decisão', *Instituto de Ciências Exatas e Informática*.

OpServices (2017) *O QUE É O GARTNER E QUAL SUA IMPORTÂNCIA PARA A TI?*, OpServices.

Oracle (2019) *Oracle Data Integrator*, Oracle.

Pereira, D. (2016) *O que é o Business Model Canvas*, *O Analista de Modelo de Negócio*.

Piton, R. (2017) *Data Warehouse – O que é Tabela Fato (e seus tipos)*.

Power, D. J. (2007) *A Brief History of Decision Support Systems*, *DSS Resources*.

Qlik Sense (2019) *Qlik Sense*, *Qlik Sense*.

Rangarajan, S. (2016) *Data Warehouse Design – Inmon versus Kimball*, *TDAN.com*.

Reis, M. (2019) *Saiba o que significa neutrófilos altos ou baixos*, *Tua Saúde*.

ReQtest (2018) *Requirements Analysis – Understanding the basics*.

ResearchGate.net (2017) *Michael Porter's value chain*, *ResearchGate.net*.

Rosa, P. (2018) *O que é a Hemoglobina globular média (HGM) e quais os valores normais*.

Santos, M. Dos (2008) *Escala fundamental de Saaty*, *ResearchGate.net*.

SaúdeBusiness (2016) *Business Intelligence na Saúde: 3 aspectos permitidos pela tecnologia*.

Significados (2018) *Hemácias: o que são, funções e características dos eritrocitos*.

SNS (2016) *Sangue | Doação*.

SNS (2019) *Dia Nacional do dador de Sangue 2019*.

SPHA, S. P. de H. (2019) *HIPERTENSÃO ARTERIAL (HTA): O QUE É?*

Stuppiello, B. (2018) *Ácido fólico: para que serve e como tomar*.

Tableau (2019a) *Business intelligence for your people*, Tableau.

Tableau (2019b) *Self-service analytics in the cloud with Tableau Online*, Tableau.

Talend (2019) *What is ETL (Extract, Transform, Load)?*, Talend.

TechDifferences (2016) *Difference Between OLTP and OLAP*, TechDifferences.

Teixeira, J. B., Filho, M. M. dos S. and Costa, C. A. D. (2018) 'Business Intelligence para uma análise da qualidade da entrega dos objetos postais', *Revista de Engenharia e Pesquisa Aplicada*, 3(3).

Unidos pela Hemofilia (2019) *O que é o sangue?*

UOL (2018) *Composição do sangue - plasma, hemácias, leucócitos e plaquetas*.

Vogt, F. (2013) *Os benefícios das ferramentas de Business Intelligence na área de saúde, Healthcare*.

Wisdom Jobs (2019) *INSIDE THE LOGICAL DATA MAP*.

Zhao, S. (2017) *What is ETL? (Extract, Transform, Load)*, Experian.

A1 Anexo 1 – Logical Data Map (IPS – Staging Area)

| Target | | | | | Source | | | | Transformation |
|------------|---------------|---------------|------------|----------|---------------|----------------|----------------|-------------|----------------|
| Table Name | Column Name | Data type | Table Type | SCD Type | DataBase Name | Table Name | Column Name | Data Type | |
| ANALISES | Data | datetime | Staging | | IPS | Analises Total | Data | Data/Hora | |
| ANALISES | Dador | numeric(18,0) | Staging | | IPS | Analises Total | Dador | Número | |
| ANALISES | Ano_Colh | numeric(18,0) | Staging | | IPS | Analises Total | Ano Colh | Número | |
| ANALISES | Num_Colh | numeric(18,0) | Staging | | IPS | Analises Total | Num Colh | Número | |
| ANALISES | ALT | varchar(255) | Staging | | IPS | Analises Total | ALT | Texto Breve | |
| ANALISES | ALTN | numeric(18,0) | Staging | | IPS | Analises Total | ALTN | Número | |
| ANALISES | Ag_HBs | varchar(255) | Staging | | IPS | Analises Total | Ag HBs | Texto Breve | |
| ANALISES | Ag_HBsN | numeric(18,0) | Staging | | IPS | Analises Total | Ag HBsN | Número | |
| ANALISES | Anti_HCV | varchar(255) | Staging | | IPS | Analises Total | Anti-HCV | Texto Breve | |
| ANALISES | Anti_HCVN | numeric(18,0) | Staging | | IPS | Analises Total | Anti-HCVN | Número | |
| ANALISES | Anti_HBc | varchar(255) | Staging | | IPS | Analises Total | Anti-HBc | Texto Breve | |
| ANALISES | Anti_HBcN | numeric(18,0) | Staging | | IPS | Analises Total | Anti-HBcN | Número | |
| ANALISES | Anti_HIV1_2 | varchar(255) | Staging | | IPS | Analises Total | Anti-HIV 1/2 | Texto Breve | |
| ANALISES | Anti_HIV1_2N | numeric(18,0) | Staging | | IPS | Analises Total | Anti-HIV 1/2N | Número | |
| ANALISES | Anti_HTLV1_2 | varchar(255) | Staging | | IPS | Analises Total | Anti-HTLV 1/2 | Texto Breve | |
| ANALISES | Anti_HTLV1_2N | numeric(18,0) | Staging | | IPS | Analises Total | Anti-HTLV 1/2N | Número | |
| ANALISES | VDRL | varchar(255) | Staging | | IPS | Analises Total | VDRL | Texto Breve | |
| ANALISES | VDRLN | varchar(255) | Staging | | IPS | Analises Total | VDRLN | Texto Breve | |

| | | | | | | | | | |
|---------------|------------|---------------|---------|--|-----|----------------|------------|-------------|--------------------------------|
| ANALISES | WBC | varchar(255) | Staging | | IPS | Analises Total | WBC | Texto Breve | |
| ANALISES | WBCN | numeric(18,0) | Staging | | IPS | Analises Total | WBCN | Número | |
| ANALISES | RBC | varchar(255) | Staging | | IPS | Analises Total | RBC | Texto Breve | |
| ANALISES | RBCN | numeric(18,0) | Staging | | IPS | Analises Total | RBCN | Número | |
| ANALISES | HGB | varchar(255) | Staging | | IPS | Analises Total | HGB | Texto Breve | |
| ANALISES | HGBN | numeric(18,0) | Staging | | IPS | Analises Total | HGBN | Número | |
| ANALISES | HCT | varchar(255) | Staging | | IPS | Analises Total | HCT | Texto Breve | |
| ANALISES | HCTN | numeric(18,0) | Staging | | IPS | Analises Total | HCTN | Número | |
| ANALISES | MCV | varchar(255) | Staging | | IPS | Analises Total | MCV | Texto Breve | |
| ANALISES | MCVN | numeric(18,0) | Staging | | IPS | Analises Total | MCVN | Número | |
| ANALISES | MCH | varchar(255) | Staging | | IPS | Analises Total | MCH | Texto Breve | |
| ANALISES | MCHN | numeric(18,0) | Staging | | IPS | Analises Total | MCHN | Número | |
| ANALISES | MCHC | varchar(255) | Staging | | IPS | Analises Total | MCHC | Texto Breve | |
| ANALISES | MCHCN | numeric(18,0) | Staging | | IPS | Analises Total | MCHCN | Número | |
| ANALISES | PLT | varchar(255) | Staging | | IPS | Analises Total | PLT | Texto Breve | |
| ANALISES | PLTN | numeric(18,0) | Staging | | IPS | Analises Total | PLTN | Número | |
| ANALISES | NEU% | varchar(255) | Staging | | IPS | Analises Total | NEU% | Texto Breve | |
| ANALISES | NEU%N | numeric(18,0) | Staging | | IPS | Analises Total | NEU%N | Número | |
| BRIGADAS | CodBrigada | varchar(255) | Staging | | IPS | Brigadas | CodBrigada | Texto Breve | |
| BRIGADAS | DesBrigada | varchar(255) | Staging | | IPS | Brigadas | DesBrigada | Texto Breve | |
| CODIGO_POSTAL | CodPostal | varchar(255) | Staging | | IPS | Codigo Postal | CodPostal | Texto Breve | PATINDEX('%[0-9]%', codpostal) |

| | | | | | | | | | |
|---------------|-----------------|---------------|---------|--|-----|-----------------|----------------------|-------------|--|
| CODIGO_POSTAL | LocPostal | varchar(255) | Staging | | IPS | Codigo Postal | LocPostal | Texto Breve | UPPER(PATINDEX('%[a-zA-Z]%', codpostal)) |
| COLHEITAS | Data | datetime | Staging | | IPS | Colheitas Total | Data | Data/Hora | |
| COLHEITAS | Externa | varchar(255) | Staging | | IPS | Colheitas Total | Externa | Texto Breve | |
| COLHEITAS | Dador | numeric(18,0) | Staging | | IPS | Colheitas Total | Dador | Número | |
| COLHEITAS | Ano_Colh | numeric(18,0) | Staging | | IPS | Colheitas Total | Ano Colh | Número | |
| COLHEITAS | Num_Colh | numeric(18,0) | Staging | | IPS | Colheitas Total | Num Colh | Número | |
| COLHEITAS | Comp_Colh | varchar(255) | Staging | | IPS | Colheitas Total | Comp Colh | Texto Breve | |
| COLHEITAS | Tip_Colh | varchar(255) | Staging | | IPS | Colheitas Total | Tip Colh | Texto Breve | |
| COLHEITAS | Abo | varchar(255) | Staging | | IPS | Colheitas Total | Abo | Texto Breve | |
| COLHEITAS | Rh | numeric(10,0) | Staging | | IPS | Colheitas Total | Rh | Texto Breve | |
| COLHEITAS | Inut | numeric(18,0) | Staging | | IPS | Colheitas Total | Inut | Número | |
| COLHEITAS | Hospital | numeric(18,0) | Staging | | IPS | Colheitas Total | Hospital | Número | |
| COLHEITAS | Devolucao1 | varchar(255) | Staging | | IPS | Colheitas Total | Devolucao1 | Texto Breve | |
| COLHEITAS | Devolucao2 | varchar(255) | Staging | | IPS | Colheitas Total | Devolucao2 | Texto Breve | |
| COLHEITAS | Devolucao3 | varchar(255) | Staging | | IPS | Colheitas Total | Devolucao3 | Texto Breve | |
| CONCLUSOES | CodigoConclusao | varchar(255) | Staging | | IPS | Conclusões | Codigo Conclusão | Texto Breve | |
| CONCLUSOES | DesConclusao | varchar(255) | Staging | | IPS | Conclusões | Designação Conclusão | Texto Breve | |
| DADOR | Data | datetime | Staging | | IPS | Dadores Total | Data | Data/Hora | |
| DADOR | Dador | numeric(18,0) | Staging | | IPS | Dadores Total | Dador | Número | |
| DADOR | Local | varchar(255) | Staging | | IPS | Dadores Total | Local | Texto Breve | |
| DADOR | Dadiva | datetime | Staging | | IPS | Dadores Total | Dadiva | Texto Breve | |
| DADOR | Ano_Colh | numeric(18,0) | Staging | | IPS | Dadores Total | Ano Colh | Número | |

| | | | | | | | | | |
|-------|-------------------|---------------|---------|--|-----|---------------|-------------------|-------------|--|
| DADOR | Num_Colh | numeric(18,0) | Staging | | IPS | Dadores Total | Num Colh | Número | |
| DADOR | Sexo | numeric(18,0) | Staging | | IPS | Dadores Total | Sexo | Número | |
| DADOR | Data_Nasc | datetime | Staging | | IPS | Dadores Total | D Nasc | Data/Hora | |
| DADOR | Est_Civil | numeric(18,0) | Staging | | IPS | Dadores Total | Est Civil | Número | |
| DADOR | Profissao | varchar(255) | Staging | | IPS | Dadores Total | Profissao | Texto Breve | |
| DADOR | Cod_Postal | numeric(18,0) | Staging | | IPS | Dadores Total | Cod Postal | Número | PATINDEX('%[0-9]%', codpostal) |
| DADOR | Loc_Postal | varchar(255) | Staging | | IPS | Dadores Total | Loc Postal | Texto Breve | UPPER(PATINDEX('%[a-zA-Z]%', codpostal)) |
| DADOR | Tot_Dadivas | numeric(18,0) | Staging | | IPS | Dadores Total | Tot Dadivas | Número | |
| DADOR | Total_Dadivas_IPS | numeric(18,0) | Staging | | IPS | Dadores Total | Total Dadivas IPS | Número | |
| DADOR | Hora_Inscricao | datetime | Staging | | IPS | Dadores Total | Hora Inscricao | Data/Hora | |
| DADOR | Hora_Triagem | datetime | Staging | | IPS | Dadores Total | Hora Triagem | Data/Hora | |
| DADOR | Peso | numeric(18,0) | Staging | | IPS | Dadores Total | Peso | Número | |
| DADOR | Altura | numeric(18,0) | Staging | | IPS | Dadores Total | Altura | Número | |
| DADOR | Tensao_Max | numeric(18,0) | Staging | | IPS | Dadores Total | Tensao Max | Número | |
| DADOR | Tensao_Min | numeric(18,0) | Staging | | IPS | Dadores Total | Tensao Min | Número | |
| DADOR | Hemoglobina | varchar(255) | Staging | | IPS | Dadores Total | Hemoglobina | Texto Breve | |
| DADOR | Abo | varchar(255) | Staging | | IPS | Dadores Total | Abo | Texto Breve | |
| DADOR | Rh | varchar(255) | Staging | | IPS | Dadores Total | Rh | Texto Breve | |
| DADOR | Trcl | varchar(255) | Staging | | IPS | Dadores Total | Trcl | Texto Breve | |
| DADOR | Colh | varchar(255) | Staging | | IPS | Dadores Total | Colh | Texto Breve | |
| DADOR | Lab | varchar(255) | Staging | | IPS | Dadores Total | Lab | Texto Breve | |

| | | | | | | | | | |
|---------------|---------------------|---------------|---------|--|-----|-----------------------------|---------------------------|-------------|--|
| ESTADO_CIVIL | Cod_Estado_Civil | varchar(255) | Staging | | IPS | Estado Civil | CodEstadoCivil | Texto Breve | |
| ESTADO_CIVIL | DesEstadoCivil | varchar(255) | Staging | | IPS | Estado Civil | DesEstadoCivil | Texto Breve | |
| HOSPITAIS | CodHospital | numeric(18,0) | Staging | | IPS | Hospitais | CodHospital | Número | |
| HOSPITAIS | DesHospital | varchar(255) | Staging | | IPS | Hospitais | DesHospital | Texto Breve | |
| PROFISSOES | CodProfissao | varchar(255) | Staging | | IPS | Profissões | CodProfissao | Texto Breve | |
| PROFISSOES | DesProfissao | varchar(255) | Staging | | IPS | Profissões | DesProfissao | Texto Breve | |
| COMPONENTES | Comp Colh | varchar(255) | Staging | | IPS | Tab Aux - Componenetes | Comp Colh | Texto Breve | |
| TIPO_DADIVA | Tipo_Dadiva | varchar(255) | Staging | | IPS | Tab Aux - Tipo de Dádiva | Código de dádiva | Texto Breve | |
| TIPO_DADIVA | DesTipoDadiva | varchar(255) | Staging | | IPS | Tab Aux - Tipo de Dádiva | designacao de dádiva | Texto Breve | |
| SEXO | codigo | numeric(18,0) | Staging | | IPS | Sexo | código | Número | |
| SEXO | sexo | varchar(255) | Staging | | IPS | sexo | sexo | Texto Breve | |
| INUTILIZACOES | Codigo_Inutilizacao | numeric(18,0) | Staging | | IPS | Inutilizações | Código Inutilização | Número | |
| INUTILIZACOES | Des_Inutilizacao | varchar(255) | Staging | | IPS | Inutilizações | Descrição Inutilização | Texto Breve | |

A2 Anexo 2 – Logical Data Map (Staging Area – Data Mart)

| Target | | | | | Source | | | | |
|--------------|-------------|---------------|------------|----------|---------------|------------|-------------|---------------|---|
| Table Name | Column Name | Data type | Table Type | SCD Type | DataBase Name | Table Name | Column Name | Data Type | Transformation |
| DimCOLHEITAS | ColheitaKey | int | Dimensi on | 1 | | | | | Surrogate key. |
| DimCOLHEITAS | Data | datetime | Dimensi on | 2 | StagingArea | COLHEITAS | Data | datetime | select Data from stagingArea.COLHEITAS |
| DimCOLHEITAS | Externa | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Externa | varchar(255) | select Externa from stagingArea.COLHEITAS; |
| DimCOLHEITAS | Dador | numeric(18,0) | Dimensi on | 2 | StagingArea | COLHEITAS | Dador | numeric(18,0) | select Dador from stagingArea.COLHEITAS; |
| DimCOLHEITAS | Ano_Colh | numeric(18,0) | Dimensi on | 2 | StagingArea | COLHEITAS | Ano_Colh | numeric(18,0) | select Ano_Colh from stagingArea.COLHEITAS |
| DimCOLHEITAS | Num_Colh | numeric(18,0) | Dimensi on | 2 | StagingArea | COLHEITAS | Num_Colh | numeric(18,0) | select Num_Colh from stagingArea.COLHEITAS |
| DimCOLHEITAS | Comp_Colh | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Comp_Colh | varchar(255) | select Comp_Colh from stagingArea.COLHEITAS |
| DimCOLHEITAS | Tip_Colh | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Tip_Colh | varchar(255) | select Tip_Colh from stagingArea.COLHEITAS |
| DimCOLHEITAS | Abo | varchar(255) | Dimensi on | 1 | StagingArea | COLHEITAS | Abo | varchar(255) | select Abo from stagingArea.COLHEITAS |
| DimCOLHEITAS | Rh | numeric(10,0) | Dimensi on | 1 | StagingArea | COLHEITAS | Rh | numeric(10,0) | select Rh from stagingArea.COLHEITAS |
| DimCOLHEITAS | Inut | numeric(18,0) | Dimensi on | 2 | StagingArea | COLHEITAS | Inut | numeric(18,0) | select a.Inu from stagingArea.COLHEITAS a, stagingArea.INUTILIZACOES b where a.INUT=b.Codigo_Inutilizacao |
| DimCOLHEITAS | Hospital | numeric(18,0) | Dimensi on | 2 | StagingArea | COLHEITAS | Hospital | numeric(18,0) | select a.Hospital from stagingArea.COLHEITAS a, |

| | | | | | | | | | |
|--------------|------------|---------------|------------|---|-------------|--------------|------------|---------------|--|
| | | | | | | | | | stagingArea.HOSPITAIS b WHERE a.Hospital=b.CodHospital |
| DimCOLHEITAS | Devolucao1 | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Devolucao1 | varchar(255) | select Devolucao1 from stagingArea.COLHEITAS |
| DimCOLHEITAS | Devolucao2 | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Devolucao2 | varchar(255) | select Devolucao2 from stagingArea.COLHEITAS |
| DimCOLHEITAS | Devolucao3 | varchar(255) | Dimensi on | 2 | StagingArea | COLHEITAS | Devolucao3 | varchar(255) | select Devolucao3 from stagingArea.COLHEITAS |
| DimDADOR | DadorKey | int | Dimensi on | 1 | | | | | Surrogate key. |
| DimDADOR | Data | datetime | Dimensi on | 2 | StagingArea | DADOR | Data | datetime | select Data from stagingArea.DADOR |
| DimDADOR | Dador | numeric(18,0) | Dimensi on | 2 | StagingArea | DADOR | Dador | numeric(18,0) | select Dador from stagingArea.DADOR |
| DimDADOR | Local | varchar(255) | Dimensi on | 2 | StagingArea | DADOR | Local | varchar(255) | select Local from stagingArea.DADOR |
| DimDADOR | Dadiva | datetime | Dimensi on | 2 | StagingArea | DADOR | Dadiva | datetime | select Dadiva from stagingArea.DADOR |
| DimDADOR | Ano_Colh | numeric(18,0) | Dimensi on | 2 | StagingArea | DADOR | Ano_Colh | numeric(18,0) | select Ano_Colh from stagingArea.DADOR |
| DimDADOR | Num_Colh | numeric(18,0) | Dimensi on | 2 | StagingArea | DADOR | Num_Colh | numeric(18,0) | select Num_Colh from stagingArea.DADOR |
| DimDADOR | Sexo | varchar(255) | Dimensi on | 2 | StagingArea | SEXO | Sexo | varchar(255) | select b.sexo from stagingArea.DADOR a, stagingArea.Sexo b where a.sexo=b.codigo |
| DimDADOR | Data_Nasc | datetime | Dimensi on | 1 | StagingArea | DADOR | Data_Nasc | datetime | select Data_Nasc from stagingArea.DADOR |
| DimDADOR | Est_Civil | VARCHAR(255) | Dimensi on | 2 | StagingArea | ESTADO_CIVIL | Est_Civil | varchar(255) | select b.DesEstadoCilvi as Est_Civil from stagingArea.DADOR a, stagingArea.ESTADO_CIVIL b where a.Est_Civil=b.Cod_Estado_civil |
| DimDADOR | Profissao | varchar(255) | Dimensi on | 2 | StagingArea | DADOR | Profissao | varchar(255) | select Profissao from stagingArea.DADOR |
| DimDADOR | Cod_Postal | numeric(18,0) | Dimensi on | 2 | StagingArea | DADOR | Cod_Postal | numeric(18,0) | select Cod_Postal from stagingArea.DADOR a, |

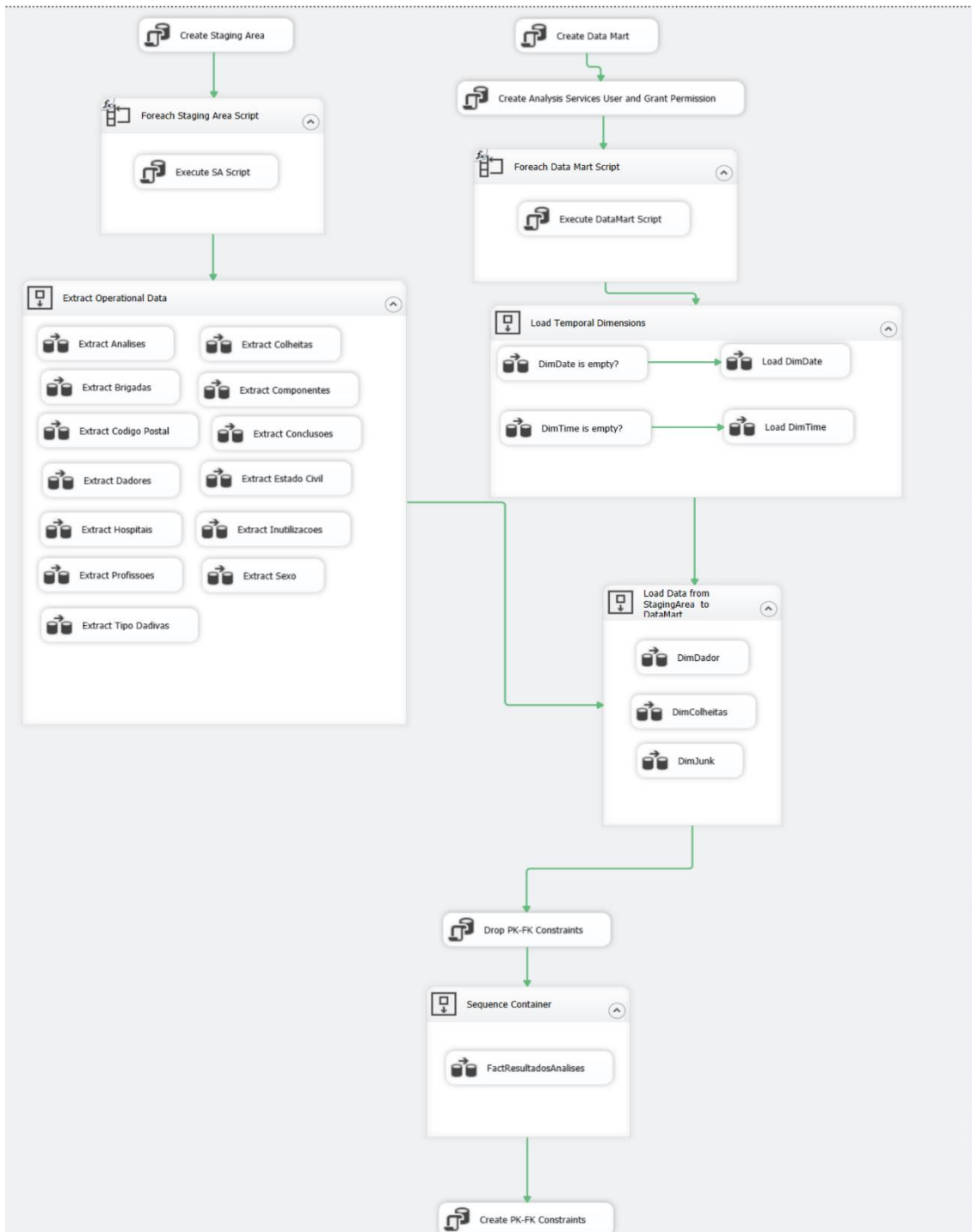
| | | | | | | | | | |
|----------|-------------------|-------------------|---------------|---|-----------------|-------|-----------------------|-------------------|--|
| | | | | | | | | | stagingArea.CODIGO_POSTAL b where a.Cod_postal=b. CodPostal |
| DimDADOR | Loc_Postal | varchar(255) | Dimensi on | 2 | StagingAre a | DADOR | Loc_Postal | varchar(255) | select Loc_Postal from stagingArea.DADOR a, stagingArea.CODIGO_POSTAL b where a.Loc_Postal=b. LocPostal |
| DimDADOR | Tot_Dadivas | numeric(18, 0) | Dimensi on | 2 | StagingAre a | DADOR | Tot_Dadivas | numeric(18, 0) | select Tot_Dadivas from stagingArea.DADOR |
| DimDADOR | Total_Dadivas_IPS | numeric(18, 0) | Dimensi on | 2 | StagingAre a | DADOR | Total_Dadivas_ IPS | numeric(18, 0) | select Total_Dadivas_IPS from stagingArea.DADOR |
| DimDADOR | Hora_Inscricao | datetime | Dimensi on | 2 | StagingAre a | DADOR | Hora_Inscricao | datetime | select Hora_Inscricao from stagingArea.DADOR |
| DimDADOR | Hora_Triagem | datetime | Dimensi on | 2 | StagingAre a | DADOR | Hora_Triagem | datetime | select Hora_Triagem from stagingArea.DADOR |
| DimDADOR | Peso | numeric(18, 0) | Dimensi on | 2 | StagingAre a | DADOR | Peso | numeric(18, 0) | select Peso from stagingArea.DADOR |
| DimDADOR | Altura | numeric(18, 0) | Dimensi on | 1 | StagingAre a | DADOR | Altura | numeric(18, 0) | select altura from stagingArea.DADOR |
| DimDADOR | Tensao_Max | numeric(18, 0) | Dimensi on | 2 | StagingAre a | DADOR | Tensao_Max | numeric(18, 0) | select Tensao_Max from stagingArea.DADOR |
| DimDADOR | Tensao_Min | numeric(18, 0) | Dimensi on | 2 | StagingAre a | DADOR | Tensao_Min | numeric(18, 0) | select Tensao_Min from stagingArea.DADOR |
| DimDADOR | Hemoglobina | varchar(255) | Dimensi on | 2 | StagingAre a | DADOR | Hemoglobina | varchar(255) | select Hemoglobina from stagingArea.DADOR |
| DimDADOR | Abo | varchar(255) | Dimensi on | 1 | StagingAre a | DADOR | Abo | varchar(255) | select Abo from stagingArea.DADOR |
| DimDADOR | Rh | varchar(255) | Dimensi on | 1 | StagingAre a | DADOR | Rh | varchar(255) | select Rh from stagingArea.DADOR |
| DimDADOR | Trcl | varchar(255) | Dimensi on | 2 | StagingAre a | DADOR | Trcl | varchar(255) | select Trcl from stagingArea.DADOR |
| DimDADOR | Colh | varchar(255) | Dimensi on | 2 | StagingAre a | DADOR | Colh | varchar(255) | select Colh from stagingArea.DADOR |
| DimDADOR | Lab | varchar(255) | Dimensi on | 2 | StagingAre a | DADOR | Lab | varchar(255) | select Lab from stagingArea.DADOR |
| DimDATE | DateKey | int | Dimensi on | 1 | | | | | |

| | | | | | | | | | |
|---------|------------------|--------------|-----------|---|-------------|----------|----------|--------------|--|
| DimDATE | year | tinyint | Dimension | 1 | | | | | |
| DimDATE | semester | tinyint | Dimension | 1 | | | | | |
| DimDATE | month | tinyint | Dimension | 1 | | | | | |
| DimDATE | monthName | nvarchar(10) | Dimension | 1 | | | | | |
| DimDATE | week | tinyint | Dimension | 1 | | | | | |
| DimDATE | DayNumberOfYear | int | Dimension | 1 | | | | | |
| DimDATE | DayNumberOfMonth | tinyint | Dimension | 1 | | | | | |
| DimDATE | DayNumberOfWeek | tinyint | Dimension | 1 | | | | | |
| DimDATE | DayOfWeek | nvarchar(10) | Dimension | 1 | | | | | |
| DimDATE | weekend | nvarchar(3) | Dimension | 1 | | | | | |
| DimDATE | trimester | tinyint | Dimension | 1 | | | | | |
| DimDATE | season | varchar(20) | Dimension | 1 | | | | | |
| DimDATE | lastDayOfMonth | nvarchar(3) | Dimension | 1 | | | | | |
| DimJunk | junkkey | int | Dimension | 1 | | | | | |
| DimJunk | ALT | varchar(255) | Dimension | 2 | StagingArea | ANALISES | ALT | varchar(255) | |
| DimJunk | Ag_HBs | varchar(255) | Dimension | 2 | StagingArea | ANALISES | Ag_HBs | varchar(255) | |
| DimJunk | Anti_HCV | varchar(255) | Dimension | 2 | StagingArea | ANALISES | Anti_HCV | varchar(255) | |
| DimJunk | Anti_HBc | varchar(255) | Dimension | 2 | StagingArea | ANALISES | Anti_HBc | varchar(255) | |

| | | | | | | | | | |
|-----------------------------|--------------|--------------|---------------|-----|------------------------------|-----------------|--------------|------------------|---|
| DimJunk | Anti_HIV1_2 | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | Anti_HIV1_2 | varchar(255) | |
| DimJunk | Anti_HTLV1_2 | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | Anti_HTLV1_2 | varchar(255) | |
| DimJunk | VDRL | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | VDRL | varchar(255) | |
| DimJunk | WBC | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | WBC | varchar(255) | |
| DimJunk | RBC | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | RBC | varchar(255) | |
| DimJunk | HGB | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | HGB | varchar(255) | |
| DimJunk | HCT | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | HCT | varchar(255) | |
| DimJunk | MCV | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | MCV | varchar(255) | |
| DimJunk | MCH | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | MCH | varchar(255) | |
| DimJunk | MCHC | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | MCHC | varchar(255) | |
| DimJunk | PLT | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | PLT | varchar(255) | |
| DimJunk | NEU% | varchar(255) | Dimensi on | 2 | StagingAre a | ANALISES | NEU% | varchar(255) | |
| FactsResultadosAnal ises | DadorKey | int | Facts | N/A | StagingAre a, DataMart | DimDADOR | DadorKey | int | where stagingArea.DADOR.Dador = DataMart.DimDADOR.Dador |
| FactsResultadosAnal ises | ColheitaKey | int | Facts | N/A | StagingAre a, DataMart | DimCOLHEI TA | ColheitaKey | int | where stagingArea.COLHEITAS.Num_Colh = DataMart.DimCOLHEITAS.Num_Colh |
| FactsResultadosAnal ises | junkkey | int | Facts | N/A | StagingAre a, DataMart | | | | |
| FactsResultadosAnal ises | Datekey | int | Facts | N/A | StagingAre a, DataMart | DimDATE | DateKey | int | where stagingArea.DADOR.Data = DataMart.DimDate.fullDate |

| | | | | | | | | | |
|-------------------------|---------------|---------------|-------|-----|-------------|----------|---------------|---------------|--|
| FactsResultadosAnalises | ALTN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | ALTN | numeric(18,0) | select ALTN from stagingArea.ANALISES |
| FactsResultadosAnalises | Ag_HBsN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | Ag_HBsN | numeric(18,0) | select Ag_HBsN from stagingArea.ANALISES |
| FactsResultadosAnalises | Anti_HCVN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | Anti_HCVN | numeric(18,0) | select Anti_HCVN from stagingArea.ANALISES |
| FactsResultadosAnalises | Anti_HBcN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | Anti_HBcN | numeric(18,0) | select Anti_HBcN from stagingArea.ANALISES |
| FactsResultadosAnalises | Anti_HIV1_2N | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | Anti_HIV1_2N | numeric(18,0) | select Anti_HIV1_2N from stagingArea.ANALISES |
| FactsResultadosAnalises | Anti_HTLV1_2N | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | Anti_HTLV1_2N | numeric(18,0) | select Anti_HTLV1_2N from stagingArea.ANALISES |
| FactsResultadosAnalises | VDRLN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | VDRLN | varchar(255) | |
| FactsResultadosAnalises | WBCN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | WBCN | numeric(18,0) | select WBCN from stagingArea.ANALISES |
| FactsResultadosAnalises | RBCN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | RBCN | numeric(18,0) | select RBCN from stagingArea.ANALISES |
| FactsResultadosAnalises | HGBN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | HGBN | numeric(18,0) | select HGBN from stagingArea.ANALISES |
| FactsResultadosAnalises | HCTN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | HCTN | numeric(18,0) | select HCTN from stagingArea.ANALISES |
| FactsResultadosAnalises | MCVN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | MCVN | numeric(18,0) | select MCVN from stagingArea.ANALISES |
| FactsResultadosAnalises | MCHN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | MCHN | numeric(18,0) | select MCHN from stagingArea.ANALISES |
| FactsResultadosAnalises | MCHCN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | MCHCN | numeric(18,0) | select MCHCN from stagingArea.ANALISES |
| FactsResultadosAnalises | PLTN | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | PLTN | numeric(18,0) | select PLTN from stagingArea.ANALISES |
| FactsResultadosAnalises | NEU%N | numeric(18,0) | Facts | N/A | StagingArea | ANALISES | NEU%N | numeric(18,0) | select NEU%N from stagingArea.ANALISES |

A3 Anexo 3 – Control flow da solução desenvolvida



A4 Anexo 4 – Visualização dos dados no Power BI

Para se visualizar os dados que se encontram no Data Mart utiliza-se a ferramenta Power BI Desktop. Esta ferramenta, após a sua instalação, permite efetuar ligação com a fonte de dados. Para tal, basta de uma forma bastante simples, seleccionar a fonte onde os dados se encontram. No Data Mart em questão, a fonte é uma base de dados “SQL Server Database “, tal como se pode visualizar na Figura 54.

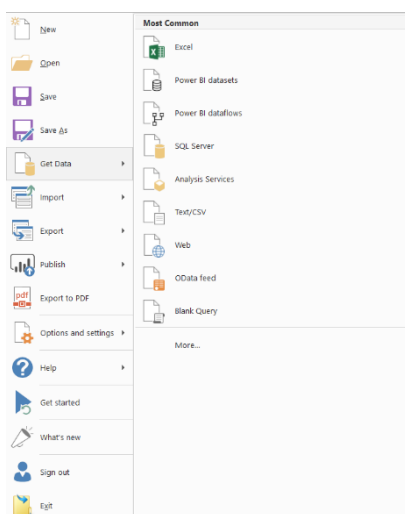


Figura 54 - Seleção da fonte de dados no Power BI

Após isso, irá surgir o pedido de autenticação à base de dados, sendo necessário colocar o username e password, caso a mesma possua esta informação.

Após aceder à base de dados, surgem todas as tabelas (Tabelas de Dimensão e de Factos), dando a oportunidade ao utilizador, de editar os dados ou apenas de carregar os mesmos. Salienta-se a importância de colocar, logo à partida, as Keys das tabelas com o formato de texto, uma vez que o Power BI assume que pode fazer cálculos com elas, caso as mesmas estejam com o formato numérico. Mesmo que os dados sejam carregados sem serem editados, esta última ação pode ser feita a qualquer momento, de acordo com as necessidades e correções que podem surgir.

A informação, já nesta ferramenta pode ser visualizada segundo três vistas (Microsoft, 2019b):

Vista de Relatório - é onde são criados os relatórios e os elementos visuais, pelo que consequentemente é a vista onde se passa maior parte do tempo.

Vista de Dados - é onde são evidenciadas as tabelas, as medidas e informações usadas no modelo de dados, associado aos relatórios. Para além disto, é onde se efetuam as transformações nos dados, caso ainda surja necessidade para tal.

Vista de Modelo – Esta vista permite ver e gerir as relações das tabelas do modelo de dados. O modelo relacional em estrela na ferramenta é exibido através da vista de Modelo, tal como se pode visualizar na Figura 55.

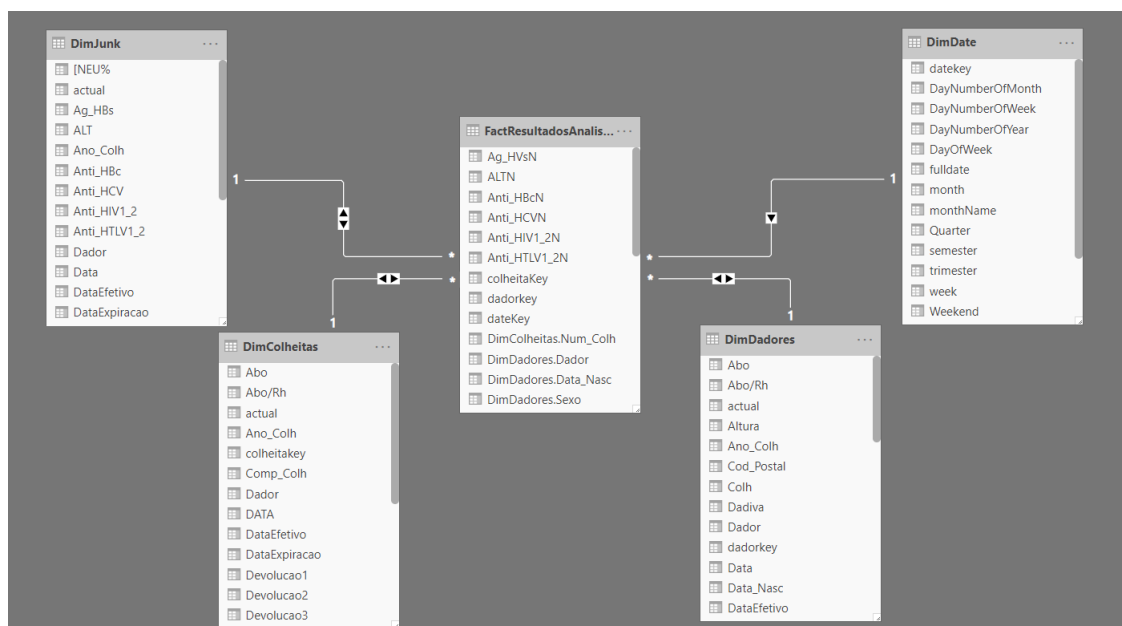



Figura 55 - Representação do modelo relacional no Power BI

Por fim, o Power BI utiliza uma linguagem denominada DAX para criar nova informação a partir dos dados que existem no modelo. Assim pode-se criar novas colunas e métricas, já que DAX consiste num “conjunto de funções, operadores e constantes que podem ser usados numa fórmula, ou expressão, para calcular e devolver um ou mais valores” (Microsoft, 2019a).

Existem também um conjunto de elementos visuais que possibilitam ao utilizador visualizar os dados, de forma a que a informação surja intuitiva, de fácil interpretação, de forma a facilitar o dia a dia dos profissionais de saúde.

Após conclusão de todo o trabalho desenvolvido no Power BI Desktop, esta ferramenta permite partilhar relatórios.

Para partilhar relatórios, basta clicar sobre . Este botão, faz com que seja aberta uma janela para se seleccionar onde se pretende publicar os relatórios. Quando se publicam os relatórios na conta do Power BI, pode-se visualizar em “A minha área de trabalho” os relatórios, tal como é evidenciado na Figura 56.

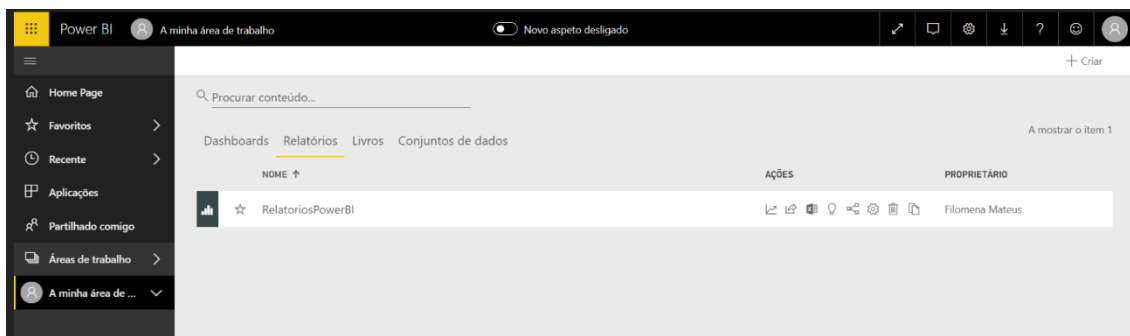


Figura 56 - Área de trabalho do Power BI

Aqui ainda se pode proceder à partilha de relatórios, clicando sobre o devido símbolo que se encontra nas “Ações”. Após isso, é evidenciado como proceder à partilha relatórios, tal como indica a Figura 57. Basta indicar o(s) email(s) com quem se pretende partilhar os relatórios, pode-se ainda incluir uma mensagem e por fim, controlar as permissões que são dadas aos utilizadores, selecionado ou não as afirmações mencionadas.

Partilhar relatório

RELATORIOSPOWERBI

Partilhar Aceder

Apenas os utilizadores com o Power BI Pro têm acesso a este relatório. Os destinatários terão o mesmo acesso a menos que a segurança de nível de linha no conjunto de dados o restrinja no futuro. [Saber mais](#)

Conceder acesso a

Introduza os endereços de e-mail

Incluir uma mensagem opcional...

- ☒ Permitir aos destinatários partilhar o relatório
- ☒ Permitir que os utilizadores criem novos conteúdos com conjuntos de dados subjacentes
- ☒ Enviar notificação por e-mail aos destinatários

Ligação do relatório ⓘ

<https://app.powerbi.com/groups/me/reports/e054f9a4-3c82-4e5a-9c11-a29abee>

Partilhar Cancelar

Figura 57 - Partilha de relatório no Power BI

É também possível visualizar todos os relatórios que foram partilhados comigo. Para tal basta clicar sobre “Partilhado Comigo” na conta do Power BI.

A5 Anexo 5 – Interações Dashboards

As figuras evidenciam alguns exemplos de interações que o utilizador pode realizar sobre as dashboards. Na primeira, Figura 58, encontram-se os filtros de Ano e Sexo Feminino selecionados e pode-se visualizar a diferença, através das barras dos gráficos, entre doadores do sexo feminino (vermelho escuro) e o total de doadores, para os mesmos anos (vermelho claro).

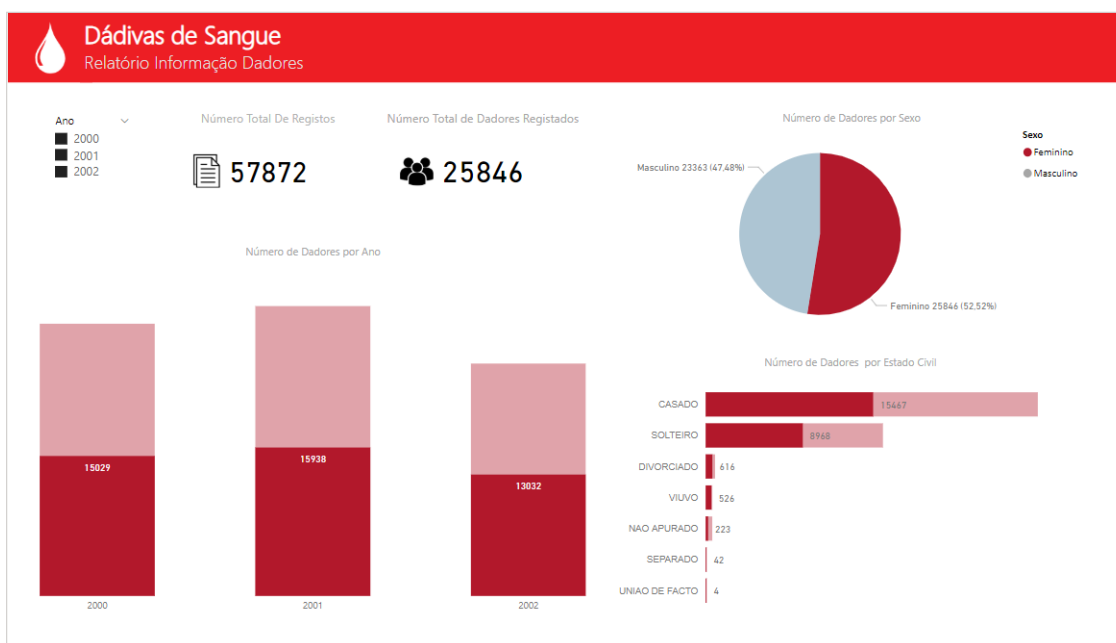


Figura 58 - Interação em dashboard “Dadores”

Na Figura 59 foi selecionado o filtro do Grupo Sanguíneo A+. Desta forma, verifica-se a diferença entre o total de colheitas e o número total de colheitas cujo grupo sanguíneo é do tipo A+.

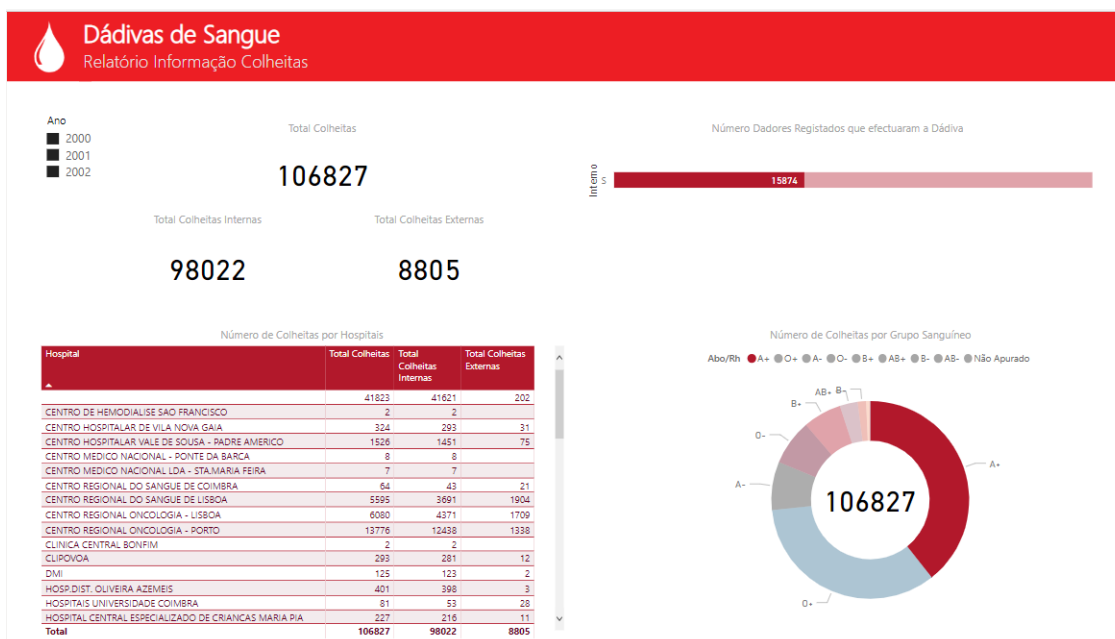


Figura 59 - Interação dashboard "Colheitas"

A Figura 60 é um exemplo de relatório, após serem seleccionados os filtros Ano e Sexo Feminino. Comparando este com o que já foi apresentado no capítulo 8, verifica-se que os valores sofreram alterações.

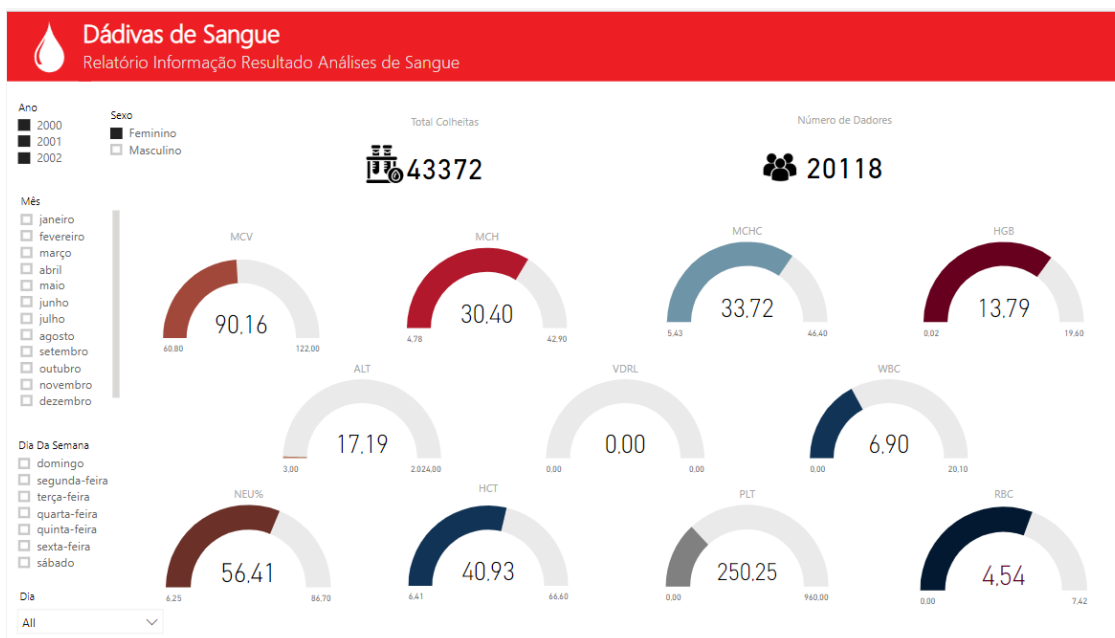


Figura 60 - Interação dashboard "Resultados Análises de Sangue"

A6 Anexo 6 – Inquérito de satisfação aos profissionais de IT

Inquérito de Satisfação à solução desenvolvida - Profissionais IT

Foi desenvolvido um sistema de Business Intelligence para ser aplicado em dados de Dádivas de Sangue e que visa auxiliar profissionais de saúde. Como forma de avaliar a solução desenvolvida, são apresentados alguns pontos que devem ser avaliados de acordo com uma escalada de 1 a 5, onde o 1 representa "Discordo Totalmente", 2 representa "Discordo", 3 representa "Não Concordo nem Discordo", 4 representa "Concordo" e 5 representa "Concordo Totalmente"

O sistema possui uma arquitetura que se adequa ao negócio.

1

2

3

4

5

☐

☐

☐

☐

☐

A arquitetura desenvolvida encontra-se preparada para novos indicadores.

1

2

3

4

5

☐

☐

☐

☐

☐

A aplicação integra facilmente com outros sistemas.

1

2

3

4

5

☐

☐

☐

☐

☐

O tempo de resposta da aplicação é adequado.

1

2

3

4

5

☐

☐

☐

☐

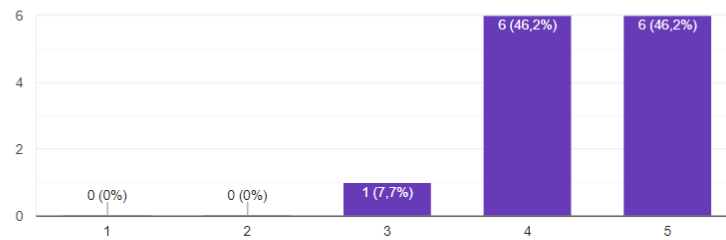
☐

SUBMITER

Figura 61 - Inquérito de Satisfação (Profissionais IT)

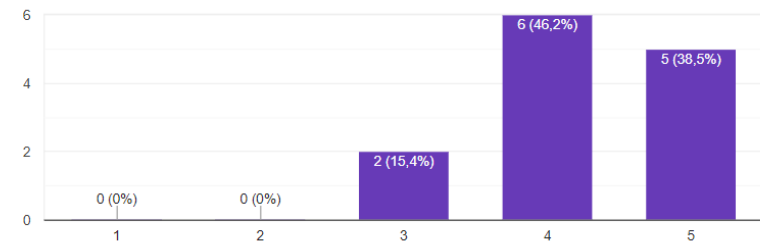
O sistema possui uma arquitetura que se adequa ao negócio.

13 respostas



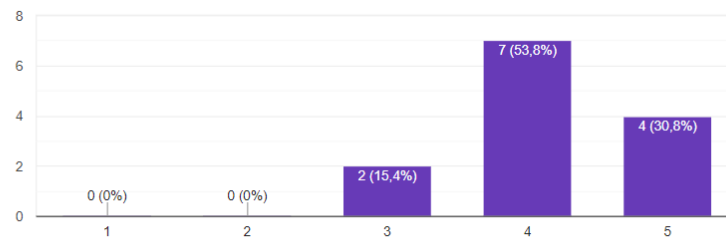
A aplicação integra facilmente com outros sistemas.

13 respostas



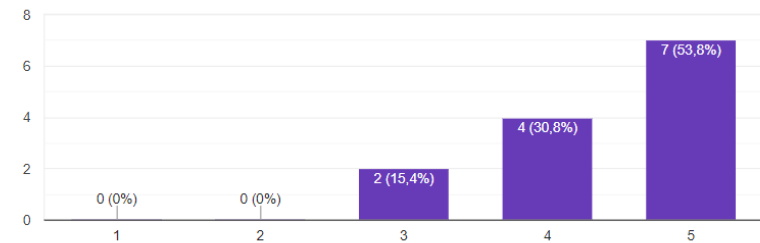
A arquitetura desenvolvida encontra-se preparada para novos indicadores.

13 respostas



O tempo de resposta da aplicação é adequado.

13 respostas



A7 Anexo 7 – Inquérito de satisfação aos profissionais de Saúde

Inquérito de Satisfação à solução desenvolvida - Profissionais de Saúde

Foi desenvolvido um sistema de Business Intelligence para ser aplicado em dados de Dídivas de Sangue e que visa auxiliar profissionais de saúde. Como forma de avaliar a solução desenvolvida, são apresentados alguns pontos que devem ser avaliados de acordo com uma escalada de 1 a 5, onde o 1 representa "Discordo Totalmente", 2 representa "Discordo", 3 representa "Não Concordo nem Discordo", 4 representa "Concordo" e 5 representa "Concordo Totalmente"

A aplicação mostra-se útil e intuitiva.

1

2

3

4

5

☐

☐

☐

☐

☐

Os filtros adequam-se ao tema em estudo.

1

2

3

4

5

☐

☐

☐

☐

☐

A informação encontra-se disponibilizada de forma a facilitar a leitura e a análise dos dados.

1

2

3

4

5

☐

☐

☐

☐

☐

O método de apresentação dos indicadores é preferível quando comparado com relatórios estáticos.

1

2

3

4

5

☐

☐

☐

☐

☐

Considera a aplicação uma boa ferramenta para exploração de dados.

1

2

3

4

5

☐

☐

☐

☐

☐

Considera que a solução desenvolvida foi conseguida de acordo com os objectivos a atingir.

1

2

3

4

5

☐

☐

☐

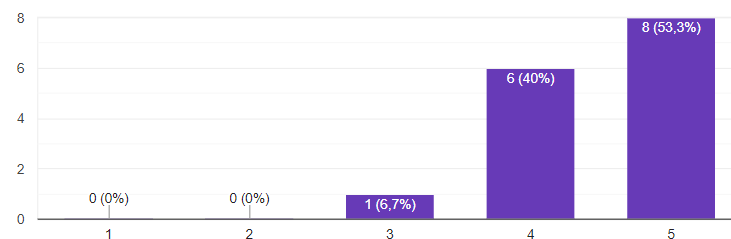
☐

☐

SUBMITER

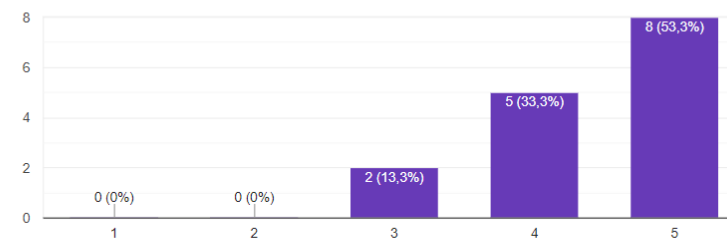
A aplicação mostra-se útil e intuitiva.

15 respostas



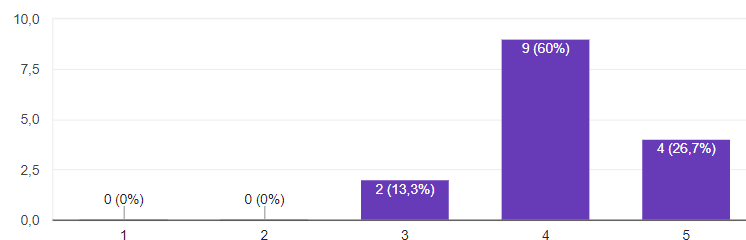
A informação encontra-se disponibilizada de forma a facilitar a leitura e a análise dos dados.

15 respostas



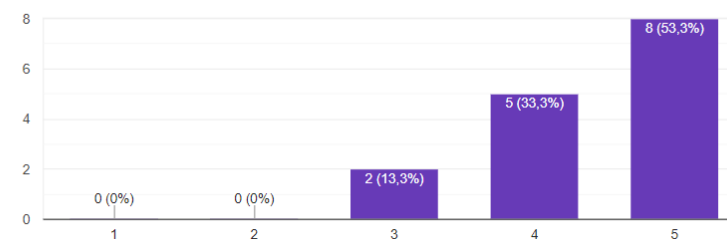
Os filtros adequam-se ao tema em estudo.

15 respostas



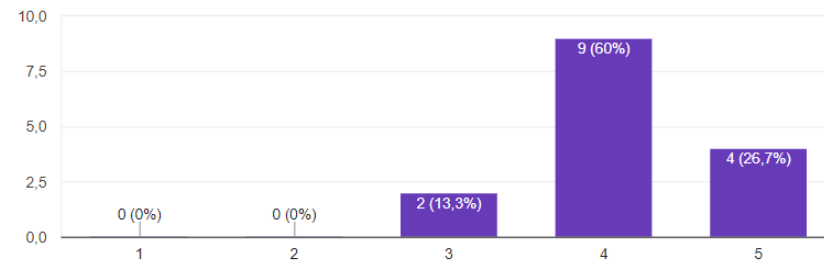
O método de apresentação dos indicadores é preferível quando comparado com relatórios estáticos.

15 respostas



Considera a aplicação uma boa ferramenta para exploração de dados.

15 respostas



Considera que a solução desenvolvida foi conseguida de acordo com os objectivos a atingir.

15 respostas

